534,869

(12)特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2004 年6 月3 日 (03.06.2004)

PCT

(10) 国際公開番号
WO 2004/047076 A1

(51) 国際特許分類[7]: G10L 15/06, G06K 9/68

(21) 国際出願番号: PCT/JP2003/014626

(22) 国際出願日: 2003 年11 月18 日 (18.11.2003)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(30) 優先権データ:
特願2002-338652
2002 年11 月21 日 (21.11.2002) JP
特願2003-89179 2003 年3 月27 日 (27.03.2003) JP
特願2003-284489 2003 年7 月31 日 (31.07.2003) JP

(71) 出願人 (米国を除く全ての指定国について): 松下電器産業株式会社 (MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.) [JP/JP]; 〒571-8501 大阪府 門真市 大字門真１００６番地 Osaka (JP).

(72) 発明者; および
(75) 発明者/出願人 (米国についてのみ): 芳澤 伸一 (YOSHIZAWA,Shinichi) [JP/JP]; 〒573-0064 大阪府 枚方市 北中振2丁目3番32-304号 Osaka (JP).

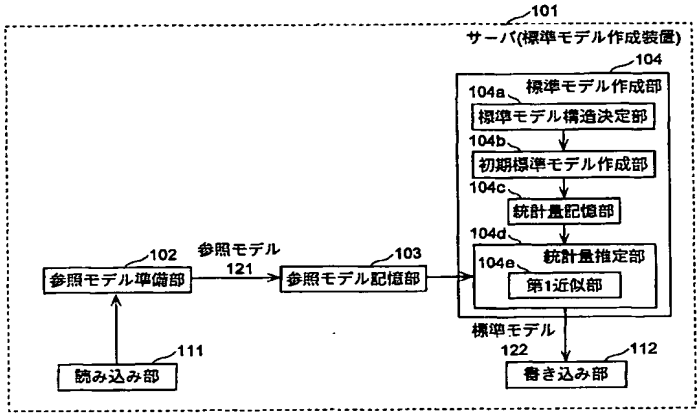
(74) 代理人: 新居 広守 (NII,Hiromori); 〒532-0011 大阪府 大阪市淀川区 西中島3丁目11番26号 新大阪末広センタービル3F 新居国際特許事務所内 Osaka (JP).

(81) 指定国 (国内): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

[続葉有]

(54) Title: STANDARD MODEL CREATING DEVICE AND STANDARD MODEL CREATING METHOD

(54) 発明の名称: 標準モデル作成装置及び標準モデル作成方法



102...REFERENCE MODEL PREPARING UNIT
111...READING UNIT
121...REFERENCE MODEL
103...REFERENCE MODEL STORAGE UNIT
101...SERVER (STANDARD MODEL CREATING DEVICE)
104...STANDARD MODEL CREATING UNIT
104a...STANDARD MODEL STRUCTURE DETERMINING SECTION
104b...INITIAL STANDARD MODEL PREPARING SECTION
104c...STATISTIC QUANTITY STORAGE SECTION
104d...STATISTIC QUANTITY ESTIMATING SECTION
104e...FIRST APPROXIMATING SECTION
122...STANDARD MODEL
112...WRITING MODEL

(57) Abstract: A standard model creating device for providing a high-precision standard model used for pattern recognition such as speech recognition, character recognition, or image recognition by using a probability model based on a hidden Markov model, the Bayesian theory, linear discrimination analysis, intention interpretation using a probability model such as a Bayesian net, and data-mining performed by using a probability model. The standard model creating device comprises a reference model preparing unit (102) for preparing

[続葉有]

WO 2004/047076 A1

(84) 指定国 (広域): ARIPO 特許 (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア特許 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ特許 (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI 特許 (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:
— 国際調査報告書

one or more reference models, a reference model storage unit (103) storing a reference model (121) prepared by the reference model preparing unit (102), and a standard model creating unit (104) for creating a standard model (122) by calculating a statistic quantity of the standard model in such a way that the probability or likelihood of one or more reference models stored in the reference model storage unit (103) is maximized or locally maximized.

(57) 要約: 隠れマルコフモデル、ベイズ理論、線形判別分析などの確率モデルによる音声認識、文字認識、画像認識などのパターン認識、ベイジアンネットなどの確率モデルによる意図理解、確率モデルによるデータマイニングなどに用いる高精度な標準モデルを提供する標準モデル作成装置であって、１以上の参照モデルを準備する参照モデル準備部（１０２）と、参照モデル準備部（１０２）が準備した参照モデル（１２１）を記憶する参照モデル記憶部（１０３）と、参照モデル記憶部（１０３）が記憶している１以上の参照モデルに対する確率又は尤度を最大化又は極大化するように標準モデルの統計量を計算して標準モデル（１２２）を作成する標準モデル作成部（１０４）とを備える。

明　細　書

標準モデル作成装置及び標準モデル作成方法

技術分野

本発明は、隠れマルコフモデル、ベイズ理論、線形判別分析などの確率モデルによる音声認識、文字認識、画像認識などのパターン認識、ベイジアンネットなどの確率モデルによる意図理解（意図の認識）、確率モデルによるデータマイニング（データ特性の認識）、確率モデルによる人物検出、指紋認証、顔認証、虹彩認証（対象を認識して特定の対象かどうかを判断する）、株価予測、天気予測などの予測（状況を認識して判断する）、複数の話者音声の合成、複数の顔画像などの合成（合成したモデルを人が認識して楽しむ）などに用いられる標準モデルの作成装置及びその方法に関する。

背景技術

近年、インターネットなどの普及により、ネットワークの大容量化、通信コストの低価格化が進んでいる。このため、ネットワークを利用することで、多くの認識用モデル（参照モデル）を収集することが可能となってきている。例えば、音声認識において、様々な研究機関で配布している多くの音声認識用モデル（子供用モデル、成人用モデル、高齢者用モデル、自動車内用モデル、携帯電話用モデルなど）をインターネットによりダウンロードすることが可能となってきている。また、ネットワークによる機器連携により、カーナビゲーションシステムなどで利用する音声認識用モデルをテレビやパソコンなどにダウンロードできるようになってきている。また、意図理解において、各地の様々な人の経験を学習した認識用モデルを、ネットワークを通して収集することが可能

となってきている。

また、認識技術の発展により、認識用モデルは、パソコン、テレビのリモコン、携帯電話、カーナビゲーションシステムなど、ＣＰＵパワー、メモリ量などの仕様の異なる幅広い機器に利用されるようになってきている。また、セキュリティーなどの認識精度が要求されるアプリケーションや、テレビのリモコンでの操作のように認識結果が出力されるまでの時間の速さが要求されるアプリケーションなど、要求仕様の異なる幅広いアプリケーションに利用されるようになってきている。

また、認識技術は、認識対象の異なる多くの環境で利用されるようになってきている。例えば、音声認識において、子供の声、成人の声、高齢者の声を認識したり、自動車内での声、携帯電話での声を認識するなど、多くの環境で利用される。

これらの社会環境の変化を鑑みると、多くの認識用モデル（参照モデル）を有効に活用することで、機器やアプリケーションの仕様、利用環境に適した精度の高い認識用モデル（標準モデル）を短時間に作成して利用者に提供することが望まれると考えられる。

音声認識などのパターン認識の分野では、認識用の標準モデルとして確率モデルを用いる方法が近年注目されており、特に、隠れマルコフモデル（以下ＨＭＭと呼ぶ）や混合ガウス分布モデル（以下ＧＭＭと呼ぶ）が広く用いられている。また、意図理解において、意図、知識、嗜好などを表す標準モデルとして確率モデルを用いる方法が近年注目されており、特に、ベイジアンネットなどが広く用いられている。また、データマイニングの分野で、データを分類するために各カテゴリの代表モデルとして確率モデルを用いる方法が注目されており、ＧＭＭなどが広く用いられている。また、音声認証、指紋認証、顔認証、虹彩認証などの認証の分野で、認証用の標準モデルとして確率モデルを用いる方法が注目

されており、ＧＭＭなどが用いられている。ＨＭＭにより表現される標準モデルの学習アルゴリズムとしてバウム・ウェルチ（Ｂａｕｍ－Ｗｅｌｃｈ）の再推定の方法が広く用いられている（例えば、今井聖著、"音声認識"、ｐｐ.１５０－１５２、共立出版株式会社、１９９５年１１月２５日発行参照）。また、ＧＭＭにより表現される標準モデルの学習アルゴリズムとしてＥＭ（Ｅｘｐｅｃｔａｔｉｏｎ－Ｍａｘｉｍｉｚａｔｉｏｎ）アルゴリズムが広く用いられている（例えば、古井貞▲ひろ▼著、"音声情報処理"、ｐｐ．１００－１０４、森北出版株式会社、１９９８年６月３０日発行参照）。ＥＭアルゴリズムでは、標準モデル

（式１）

$$\sum_{m=1}^{M_f} \omega_{f(m)} f(x;\mu_{f(m)},\sigma^2_{f(m)})$$

（ここで、

（式２）

$$f(x;\mu_{f(m)},\sigma^2_{f(m)}) \quad (m=1,2,...,M_f)$$

はガウス分布を表し、

（式３）

$$x=(x_{(1)},x_{(2)},...,x_{(J)})\in R^J$$

はＪ（≧１）次元の入力データを表す）における統計量である混合重み係数

（式４）

、Ｊ（≧１）次元の平均値

（式５）

$$\mu_{f(m)} = (\mu_{f(m,1)}, \mu_{f(m,2)}, \ldots, \mu_{f(m,J)}) \in R^J$$

$$(m = 1,2,\ldots, M_f, j = 1,2,\ldots, J)$$

及びＪ（≧１）次元の分散値（共分散行列のＪ個の対角成分）

（式６）

$$\sigma^2_{f(m)} = (\sigma^2_{f(m,1)}, \sigma^2_{f(m,2)}, \ldots, \sigma^2_{f(m,J)}) \in R^J$$

$$(m = 1,2,\ldots, M_f, j = 1,2,\ldots, J)$$

を、

Ｎ個の学習データ

（式７）

$$x[i] = (x_{(1)}[i], x_{(2)}[i], \ldots, x_{(J)}[i]) \in R^J \quad (i = 1,2,\ldots, N)$$

を用いて、学習データに対する尤度

（式８）

$$\log P = \sum_{i=1}^{N} \log[\sum_{m=1}^{M_f} \omega_{f(m)} f(x[i]; \mu_{f(m)}, \sigma^2_{f(m)})]$$

を最大化もしくは極大化するように、

（式９）

$$\omega_{f(m)} = \frac{\sum_{i=1}^{N} \gamma(x[i], m)}{\sum_{k=1}^{M_f} \sum_{i=1}^{N} \gamma(x[i], k)}$$

$$(m = 1,2,..., M_f)$$

（式１０）

$$\mu_{f(m,j)} = \frac{\sum_{i=1}^{N} \gamma(x[i], m) x_{(j)}}{\sum_{i=1}^{N} \gamma(x[i], m)}$$

$$(m = 1,2,..., M_f, j = 1,2,..., J)$$

（式１１）

$$\sigma^2_{f(m,j)} = \frac{\sum_{i=1}^{N} \gamma(x[i], m)(x_{(j)} - \mu_{f(m,j)})^2}{\sum_{i=1}^{N} \gamma(x[i], m)}$$

$$(m = 1,2,..., M_f, j = 1,2,..., J)$$

（ここで、

（式１２）

$$\gamma(x[i], m) = \frac{\omega_{f(m)} f(x[i]; \mu_{f(m)}, \sigma^2_{f(m)})}{\sum_{k=1}^{M_f} \omega_{f(k)} f(x[i]; \mu_{f(k)}, \sigma^2_{f(k)})} \qquad (m = 1,2,..., M_f)$$

である）を利用して１以上繰り返して計算して学習を行う。また、ベイズ推定法（例えば、繁桝算男著、"ベイズ統計入門"、ｐｐ.４２－５３、東京大学出版会、１９８５年４月３０日発行参照）などの方法も提案さ

れている。バウム・ウェルチの再推定の方法、ＥＭアルゴリズム、ベイズ推定法のいずれの学習アルゴリズムも、学習データに対する確率（尤度）を最大化もしくは極大化するように標準モデルのパラメータ（統計量）を計算して標準モデルを作成する。これらの学習方法では、確率（尤度）を最大化もしくは極大化するという数学的な最適化が実現されている。

上記の学習方法を音声認識の標準モデルの作成に用いた場合、多様な話者や雑音などの音響的特徴量の変動に対応するために多数の音声データで標準モデルを学習することが望ましい。また、意図理解に用いた場合、多様な話者や状況などの変動に対応するために多数のデータで標準モデルを学習することが望ましい。また、虹彩認証に用いた場合、太陽光、カメラ位置・回転などの変動に対応するために多数の虹彩画像データで標準モデルを学習することが望ましい。しかしながら、このような多量のデータを取り扱う場合、学習に膨大な時間がかかるため、利用者に標準モデルを短時間に提供できない。また、多量のデータを蓄積するためのコストが膨大となる。また、ネットワークを利用してデータを収集した場合、通信コストが膨大となる。

一方、複数のモデル（以下、標準モデルの作成のために参照用として準備されるモデルを「参照モデル」と呼ぶ。）を合成することで標準モデルを作成する方法が提案されている。参照モデルは、多くの学習データを確率分布の母数（平均、分散など）で表現した確率分布モデルであり、多くの学習データの特徴を少数のパラメータ（母数）で集約したものである。以下に示す従来技術では、モデルはガウス分布で表現されている。

第１の従来方法では、参照モデルはＧＭＭで表現されており、複数の参照モデルのＧＭＭを重み付きで合成することで標準モデルを作成している（例えば、特開平４－１２５５９９号公報に開示された技術）。

また、第２の従来方法では、第１の従来方式に加えて、学習データに対する確率（尤度）を最大化あるいは極大化して線形結合された混合重みを学習することで標準モデルを作成している（例えば、特開平１０－２６８８９３号公報に開示された技術）。

また、第３の従来方法では、標準モデルの平均値を参照モデルの平均値の線形結合で表現し、入力データに対する確率（尤度）を最大化あるいは極大化して線形結合係数を学習することで標準モデルを作成している。ここでは学習データとして特定話者の音声データを用いており標準モデルを音声認識用の話者適応モデルとして用いている（例えば、Ｍ．Ｊ．Ｆ．Ｇａｌｅｓ、"Ｃｌｕｓｔｅｒ Ａｄａｐｔｉｖｅ Ｔｒａｉｎｉｎｇ Ｆｏｒ Ｓｐｅｅｃｈ Ｒｅｃｏｇｎｉｔｉｏｎ"、１９９８年、ＩＣＳＬＰ９８予稿集、ｐｐ．１７８３－１７８６）。

また、第４の従来方法では、参照モデルは単一ガウス分布で表現されており、複数の参照モデルのガウス分布を合成したのちに、クラスタリングにより同一クラスに属するガウス分布を統合することで標準モデルを作成している（例えば、特開平９－８１１７８号公報に開示された技術）。

また、第５の従来方法では、複数の参照モデルは同数の混合数の混合ガウス分布で表現され、各ガウス分布には１対１に対応した通し番号が付与されている。標準モデルは、同一の通し番号をもつガウス分布を合成することにより作成される。合成する複数の参照モデルは利用者に音響的に近い話者で作成されたモデルであり、作成させる標準モデルは話者適応モデルである（例えば、芳澤、外６名、"十分統計量と話者距離を用いた音韻モデルの教師なし学習法"、２００２年３月１日、電子情報通信学会、Ｖｏｌ．Ｊ８５－Ｄ－ＩＩ、Ｎｏ．３、ｐｐ．３８２－３８９）。

しかしながら、第１の従来方法では、合成する参照モデル数の増加と

ともに標準モデルの混合数が増加して、標準モデルのための記憶容量、認識処理量が膨大となり実用的でない。また、仕様に応じて標準モデルの混合数を制御することができない。この課題は、合成する参照モデルの数の増加に伴い顕著になってくると考えられる。

第２の従来方法では、合成する参照モデル数の増加とともに標準モデルの混合数が増加して、標準モデルのための記憶容量、認識処理量が膨大となり実用的でない。また、仕様に応じて標準モデルの混合数を制御することができない。また、標準モデルは、参照モデルの単純な混合和であり学習するパラメータが混合重みに限定されているため、高精度の標準モデルが作成できない。また、標準モデルの作成において、多くの学習データを用いて学習を行っているため学習時間がかかる。これらの課題は、合成する参照モデルの数の増加に伴い顕著になってくると考えられる。

第３の従来方法では、学習するパラメータが参照モデルの平均値の線形結合係数に限定されているため高精度の標準モデルが作成できない。また、標準モデルの作成において、多くの学習データを用いて学習を行っているため学習時間がかかる。

第４の従来方法では、クラスタリングをヒューリスティックに行うため高精度の標準モデルを作成することが困難である。また、参照モデルは単一のガウス分布であるため精度が低く、それらを統合した標準モデルの精度は低い。認識精度に関する課題は、合成する参照モデルの数の増加に伴い顕著になってくると考えられる。

第５の従来方法では、標準モデルは、同一の通し番号をもつガウス分布を合成することにより作成されるが、最適な標準モデルを作成するためには、一般的には合成するガウス分布は１対１に対応するとは限らないため、認識精度が低下する。また、複数の参照モデルが異なる混合数

をもつ場合に標準モデルを作成することができない。また、一般的には、参照モデルにおけるガウス分布に通し番号が付与されておらず、この場合に標準モデルを作成することができない。また、仕様に応じて標準モデルの混合数を制御することができない。

発明の開示

そこで、本発明は、このような問題点に鑑みてなされたものであり、隠れマルコフモデル、ベイズ理論、線形判別分析などの確率モデルによる音声認識、文字認識、画像認識などのパターン認識、ベイジアンネットなどの確率モデルによる意図理解（意図の認識）、確率モデルによるデータマイニング（データ特性の認識）、株価予測、天気予測などの予測（状況を認識して判断する）などに用いられる高精度な標準モデルを作成する標準モデル作成装置等を提供することを目的とする。

また、本発明は、学習のためのデータや教師データを必要とすることなく、簡易に標準モデルを作成することが可能な標準モデル作成装置等を提供することをも目的とする。

さらに、本発明は、標準モデルを利用する認識の対象にふさわしい標準モデルを作成したり、標準モデルを用いて認識処理を実行する装置の仕様や環境に適した標準モデルを作成することが可能な汎用性及び柔軟性に優れた標準モデル作成装置等を提供することをも目的とする。

本発明で用いる「認識」とは、音声認識などの狭義の意味での認識だけではなく、パターンマッチング、識別、認証、ベイズ推定や予測など、確率で表現された標準モデルを利用するもの全般を意味する。

上記目的を達成するために、本発明に係る標準モデル作成装置は、事象の集合と事象または事象間の遷移の出力確率とによって定義される認識用のモデルである標準モデルを作成する装置であって、特定の対象を

認識するために予め作成されたモデルである１以上の参照モデルを記憶する参照モデル記憶手段と、前記参照モデル記憶手段に記憶された１以上の参照モデルに対する標準モデルの確率または尤度を最大化または極大化するように当該標準モデルの統計量を計算することによって標準モデルを作成する標準モデル作成手段とを備えることを特徴とする。

たとえば、音声認識用の標準モデル作成装置として、音声の特徴を示す周波数のパラメータを出力確率で表現する確率モデルを用いて、特定の属性を有する音声の特徴を示す音声認識用の標準モデルを作成する装置であって、一定の属性を有する音声の特徴を示す確率モデルである１以上の参照モデルを記憶する参照モデル記憶手段と、前記参照モデル記憶手段に格納された１以上の参照モデルの統計量を用いて前記標準モデルの統計量を計算することによって標準モデルを作成する標準モデル作成手段とを備え、前記標準モデル作成手段は、作成する標準モデルの構造を決定する標準モデル構造決定部と、構造が決定された標準モデルを特定する統計量の初期値を決定する初期標準モデル作成部と、初期値が決定された標準モデルの前記参照モデルに対する確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を推定して計算する統計量推定部とを有することを特徴とする。

これによって、１以上の参照モデルに対する標準モデルの確率又は尤度を最大化又は極大化するように標準モデルの統計量が計算され、標準モデルが作成されるので、音声データ等の学習データや教師データを必要とすることなく簡易に標準モデルが作成されるとともに、既に作成された複数の参照モデルを総合的に勘案した高精度な標準モデルが作成される。

ここで、前記標準モデル作成装置は、さらに、外部から参照モデルを取得して前記参照モデル記憶手段に格納すること、及び、参照モデルを

作成して前記参照モデル記憶手段に格納することの少なくとも一方を行う参照モデル準備手段を備えてもよい。例えば、音声認識用に適用した場合であれば、音声の特徴を示す周波数のパラメータを出力確率で表現する確率モデルを用いて、特定の属性を有する音声の特徴を示す音声認識用の標準モデルを作成する装置であって、一定の属性を有する音声の特徴を示す確率モデルである１以上の参照モデルを記憶するための参照モデル記憶手段と、外部から参照モデルを取得して前記参照モデル記憶手段に格納すること、及び、新たな参照モデルを作成して前記参照モデル記憶手段に格納することの少なくとも一方を行う参照モデル準備手段と、所定の構造をもつ当該標準モデルの統計量の初期値を準備し、前記参照モデル記憶手段に格納された１以上の参照モデルに対する標準モデルの確率又は尤度を最大化又は極大化するように、前記参照モデルの統計量を用いて当該標準モデルの統計量を計算することによって標準モデルを作成する標準モデル作成手段とを備えることを特徴とする。

これによって、標準モデル作成装置の外部から新たな参照モデルを取り込み、取り込んだ参照モデルに基づいた標準モデルの作成が可能となるので、様々な認識対象に対応した汎用性の高い標準モデル作成装置が実現される。

また、前記標準モデル作成装置は、さらに、認識の対象に関する情報である利用情報を作成する利用情報作成手段と、作成された前記利用情報に基づいて、前記参照モデル記憶手段に記憶されている参照モデルの中から１以上の参照モデルを選択する参照モデル選択手段とを備え、前記標準モデル作成手段は、前記参照モデル選択手段が選択した参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を計算してもよい。

これによって、利用者の特徴、利用者の年齢、性別、利用環境などの

利用情報に基づいて、準備された複数の参照モデルの中から認識対象に適した参照モデルだけが選択され、それら参照モデルを統合した標準モデルが作成されるので、認識対象により特化した精度の高い標準モデルが作成される。

ここで、前記標準モデル作成装置は、さらに、前記利用情報と選択された参照モデルに関する情報との類似度を算出して、前記類似度が所定のしきい値以上であるか否かを判定して判定信号を作成する類似度判定手段を備えてもよい。

これによって、利用情報にふさわしい（近い）参照モデルが参照モデル記憶手段に存在しない場合に、参照モデルの準備の要求を行うことができる。

また、前記標準モデル作成装置には、通信路を介して端末装置が接続され、前記標準モデル作成装置は、さらに、認識の対象に関する情報である利用情報を前記端末装置から受信する利用情報受信手段と、受信された前記利用情報に基づいて、前記参照モデル記憶手段に記憶されている参照モデルの中から１以上の参照モデルを選択する参照モデル選択手段とを備え、前記標準モデル作成手段は、前記参照モデル選択手段が選択した参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を計算してもよい。

これによって、通信路を介して送信されてきた利用情報に基づいて標準モデルが作成されるので、遠隔制御による標準モデルの生成が可能になるとともに、通信システムを基盤とする認識システムの構築が実現される。

また、前記標準モデル作成装置は、さらに、作成する標準モデルの仕様に関する情報である仕様情報を作成する仕様情報作成手段を備え、前記標準モデル作成手段は、前記仕様情報作成手段が作成した仕様情報に

基づいて、前記参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を計算してもよい。

これによって、標準モデルを使用する装置のＣＰＵパワー、記憶容量、要求される認識精度、要求される認識処理時間などの仕様情報に基づいて標準モデルが作成されるので、特定の仕様条件を満たす標準モデルの生成が可能となり、計算エンジン等の認識処理に必要なリソース環境に適した標準モデルの生成が実現される。

ここで、前記仕様情報は、例えば、標準モデルを使用するアプリケーションプログラムの種類に対応づけられた仕様を示すような情報であってもよい。そして、前記標準モデル作成装置は、さらに、標準モデルを使用するアプリケーションプログラムと標準モデルの仕様との対応を示すアプリケーション仕様対応データベースを前記仕様情報として保持する仕様情報保持手段を備え、前記標準モデル作成手段は、前記仕様情報保持手段に保持されたアプリケーション仕様対応データベースから、起動されるアプリケーションプログラムに対応する仕様を読み出し、読み出した仕様に基づいて、前記参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を計算してもよい。

これによって、各アプリケーションごとに対応づけられた仕様に沿って標準モデルが作成されるので、アプリケーションごとに最適な標準モデルが作成され、標準モデルが使用される認識システム等における認識精度が向上される。

また、前記標準モデル作成装置には、通信路を介して端末装置が接続され、前記標準モデル作成装置は、さらに、作成する標準モデルの仕様に関する情報である仕様情報を前記端末装置から受信する仕様情報受信手段を備え、前記標準モデル作成手段は、前記仕様情報受信手段が受信

した仕様情報に基づいて、前記参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を計算してもよい。

これによって、通信路を介して送信されてきた仕様情報に基づいて標準モデルが作成されるので、遠隔制御による標準モデルの生成が可能になるとともに、通信システムを基盤とする認識システムの構築が実現される。

たとえば、前記参照モデル及び前記標準モデルは、１以上のガウス分布を用いて表現され、前記標準モデル作成手段は、前記仕様情報に基づいて、前記標準モデルの混合分布数（ガウス分布の数）を決定してもよい。

これによって、作成される標準モデルに含まれるガウス分布の混合分布数が動的に決定されることとなり、認識処理が実行される環境や要求仕様等に応じて標準モデルの構造を制御することが可能となる。例として、標準モデルを使用する認識装置のＣＰＵパワーが小さい場合、記憶容量が小さい場合、要求される認識処理時間が短い場合などは標準モデルの混合分布数を少なく設定して仕様に合わせることができ、一方、要求される認識精度が高い場合などは混合分布数を多く設定して認識精度を高くすることができる。

なお、上記利用情報あるいは仕様情報を用いて標準モデルを作成する場合において、参照モデル準備手段は必ずしも必要ではない。たとえば、利用者の要求に基づいて、あるいは、利用者の要求とは無関係に、予め参照モデルを標準モデル作成装置内に記憶させた状態で標準モデル作成装置を出荷し、利用情報や仕様情報を用いて標準モデルを作成することが可能だからである。

また、前記参照モデル及び前記標準モデルは、１以上のガウス分布を

用いて表現され、前記参照モデル記憶手段は、少なくとも１対の参照モデルの混合分布数（ガウス分布の数）が異なる参照モデルを記憶し、前記標準モデル作成手段は、少なくとも１対の参照モデルの混合分布数（ガウス分布の数）が異なる参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を計算してもよい。

これによって、混合分布数が異なる参照モデルに基づいて標準モデルが作成されるので、予め準備された多種多様な構造の参照モデルに基づく標準モデルの作成が可能となり、より認識対象に適した精度の高い標準モデルの作成が実現される。

また、前記標準モデル作成装置は、さらに、前記標準モデル作成手段が作成した標準モデルを記憶する標準モデル記憶手段を備えてもよい。

これによって、作成された標準モデルを一時的にバッファリングしておき、送信要求に対してすぐに出力したり、他の装置に提供するデータサーバとしての役割を果たしたりすることが可能となる。

また、前記標準モデル作成装置には、通信路を介して端末装置が接続され、前記標準モデル作成装置は、さらに、前記標準モデル作成手段が作成した標準モデルを前記端末装置に送信する標準モデル送信手段を備えてもよい。

これによって、作成された標準モデルは空間的に離れた場所に置かれた外部装置に送信されるので、本標準モデル作成装置を標準モデル作成エンジンとして独立させたり、標準モデル作成装置を通信システムにおけるサーバとして機能させたりすることが可能になる。

また、前記標準モデル作成装置には、通信路を介して端末装置が接続され、前記標準モデル作成装置は、さらに、前記端末装置から送信される参照モデルを受信する参照モデル受信手段を備え、前記標準モデル作

成手段は、少なくとも前記参照モデル受信手段が受信した参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を計算してもよい。

これによって、端末装置が保持した利用環境にふさわしい参照モデルを、通信路を介して送信して、送信した参照モデルを用いて標準モデルを作成できるため、より認識対象に適した精度の高い標準モデルの作成が実現される。例として、利用者Aが環境Aで利用していた参照モデルAが端末装置に保持されており利用者Aは環境Bで利用したい場合、参照モデルAを利用して標準モデルを作成することにより、利用者Aの特徴を反映した精度の高い標準モデルを作成することができる。

また、前記参照モデル準備手段は、さらに、前記参照モデル記憶手段が記憶する参照モデルの更新及び追加の少なくとも一方を行ってもよい。たとえば、前記標準モデル作成装置には、通信路を介して端末装置が接続され、前記標準モデル作成装置は、さらに、前記端末装置から送信される参照モデルを受信する参照モデル受信手段を備え、前記参照モデル準備手段は、前記参照モデル受信手段が受信した参照モデルを用いて前記参照モデル記憶手段が記憶する参照モデルの更新及び追加の少なくとも一方を行ってもよい。

これによって、準備される参照モデルの追加、更新等が行われるので、様々な認識対象用のモデルを参照モデルとして追加したり、より精度の高い参照モデルに置き換えたりすることが可能となり、更新した参照モデルによる標準モデルの再生成や、生成された標準モデルを参照モデルとして再び標準モデルを作成するというフィードバックによる学習等が可能となる。

また、前記標準モデル作成手段は、作成する標準モデルの構造を決定する標準モデル構造決定部と、構造が決定された前記標準モデルを特定

する統計量の初期値を決定する初期標準モデル作成部と、前記参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を推定して計算する統計量推定部とを有するように構成してもよい。このとき、前記初期標準モデル作成部は、前記統計量推定部が標準モデルの統計量を計算するために用いる、１以上の前記参照モデルを用いて前記標準モデルを特定する統計量の初期値を決定してもよい。たとえば、前記初期標準モデル作成部は、標準モデルの種類を識別するクラスＩＤに基づいて、前記初期値を決定してもよい。具体的には、前記初期標準モデル作成部は、前記クラスＩＤと前記初期値と前記参照モデルとの対応を示す対応表を保持し、前記対応表に従って、前記初期値を決定してもよい。

これによって、標準モデルが使用される認識の対象の種類ごとにクラスＩＤを付与しておくことで、最終的に必要とされる標準モデルと共通の性質をもつ初期標準モデルを使用することができるので、精度の高い標準モデルが作成される。

以上のように、本発明により、隠れマルコフモデル、ベイズ理論、線形判別分析などの確率モデルによる音声認識、文字認識、画像認識などのパターン認識、ベイジアンネットなどの確率モデルによる意図理解（意図の認識）、確率モデルによるデータマイニング（データ特性の認識）、確率モデルによる人物検出、指紋認証、顔認証、虹彩認証（対象を認識して特定の対象かどうかを判断する）、株価予測、天気予測などの予測（状況を認識して判断する）などに用いる高精度な標準モデルが提供され、その実用的価値は極めて高い。

なお、本発明は、このような標準モデル作成装置として実現することができるだけでなく、標準モデル作成装置が備える特徴的な構成要素をステップとする標準モデル作成方法として実現したり、それらのステッ

プをコンピュータに実行させるプログラムとして実現したりすることができる。そして、そのプログラムをＣＤ－ＲＯＭ等の記録媒体やインターネット等の伝送媒体を介して配信することができるのは言うまでもない。

図面の簡単な説明

図１は、本発明の第１の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

図２は、同サーバの動作手順を示すフローチャートである。

図３は、図１における参照モデル記憶部に記憶されている参照モデルの例を示す図である。

図４は、図２におけるステップＳ１０１（標準モデルの作成）の詳細な手順を示すフローチャートである。

図５は、図１における第１近似部１０４ｅによる近似計算を説明する図である。

図６は、参照モデルを選択する際の画面表示例を示す図である。

図７（ａ）は、作成する標準モデルの構造（混合分布数）を指定する際の画面表示例を示し、図７（ｂ）は、仕様情報を選択する際の画面表示例を示す図である。

図８は、標準モデルを作成しているときの進捗状況を示す画面表示例を示す図である。

図９は、本発明の第２の実施の形態における標準モデル作成装置に係るＳＴＢの全体構成を示すブロック図である。

図１０は、同ＳＴＢの動作手順を示すフローチャートである。

図１１は、図１０における参照モデル記憶部に記憶されている参照モデルの例を示す図である。

図１２は、図１０における第２近似部による近似計算を説明する図である。

図１３は、本発明の第３の実施の形態における標準モデル作成装置に係るＰＤＡの全体構成を示すブロック図である。

図１４は、同ＰＤＡの動作手順を示すフローチャートである。

図１５は、図１３における参照モデル記憶部に記憶されている参照モデルの例を示す図である。

図１６は、同ＰＤＡの選択画面の一例を示す。

図１７は、図１３における統計量推定部による統計量の推定手順を示す概念図である。

図１８は、図１３における第３近似部による近似計算を説明する図である。

図１９は、本発明の第４の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

図２０は、同サーバの動作手順を示すフローチャートである。

図２１は、同サーバの動作手順を説明するための参照モデル及び標準モデルの一例を示す図である。

図２２は、利用情報としての個人情報を入力する際の画面表示例を示す図である。

図２３は、本発明の第５の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

図２４は、同サーバの動作手順を示すフローチャートである。

図２５は、同サーバの動作手順を説明するための参照モデル及び標準モデルの一例を示す図である。

図２６は、本発明の第６の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

図２７は、同サーバの動作手順を示すフローチャートである。

図２８は、同サーバの動作手順を説明するための参照モデル及び標準モデルの一例を示す図である。

図２９は、本発明の第７の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

図３０は、同サーバの動作手順を示すフローチャートである。

図３１は、同サーバの動作手順を説明するための参照モデル及び標準モデルの一例を示す図である。

図３２は、本発明の第８の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。

図３３は、携帯電話９０１の動作手順を示すフローチャートである。

図３４は、参照モデル記憶部に格納されている参照モデルの一例を示す図である。

図３５は、新たたに参照モデル記憶部に格納された参照モデルの一例を示す図である。

図３６は、利用情報を作成するときの画面表示例を示す図である。

図３７は、参照モデルを準備するときの画面表示例を示す図である。

図３８は、第３近似部を用いて作成した標準モデルを用いた認識実験の結果を示すグラフである。

図３９は、第３の実施の形態における第２近似部により作成された標準モデルによる認識実験の結果を示すグラフである。

図４０は、本発明の第９の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。

図４１は、アプリ・仕様情報対応データベースのデータ例を示す図である。

図４２は、ＰＤＡ１００１の動作手順を示すフローチャートである。

図４３は、参照モデル記憶部に格納されている参照モデルの一例を示す図である。

図４４は、初期標準モデル作成部によるクラスタリングによる初期値の決定方法を示すフローチャートである。

図４５は、図４４におけるステップＳ１００４の具体例を示す図である。

図４６は、図４４におけるステップＳ１００５の具体例を示す図である。

図４７は、図４４におけるステップＳ１００６の具体例を示す図である。

図４８は、図４４におけるステップＳ１００８の具体例を示す図である。

図４９は、本発明の第１０の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

図５０は、同サーバの動作手順を示すフローチャートである。

図５１は、本発明に係る標準モデル作成装置を具体的に適用したシステム例を示す図である。

図５２は、クラスＩＤ・初期標準モデル・参照モデル対応表の例を示す図である。

図５３は、図５２のクラスＩＤ・初期標準モデル・参照モデル対応表における参照モデル８ＡＡ～ＡＺの例を示す図である。

図５４は、図５２のクラスＩＤ・初期標準モデル・参照モデル対応表における参照モデル６４ＺＡ～ＺＺの例を示す図である。

図５５は、図５２のクラスＩＤ・初期標準モデル・参照モデル対応表における初期標準モデル８Ａ～６４Ｚの例を示す図である。

図５６は、クラスＩＤ・初期標準モデル・参照モデル対応表の作成方

法を示すフローチャートである。

図５７は、図５６におけるステップＳ１１００の具体例を示す図である。

図５８は、図５６におけるステップＳ１１０２の具体例を示す図である。

図５９は、図５６におけるステップＳ１１０３の具体例を示す図である。

図６０は、図５６におけるステップＳ１１０４の具体例を示す図である。

図６１は、端末がサーバと通信することによってクラスＩＤ・初期標準モデル・参照モデル対応表を完成させる手順を示す図である。

図６２は、クラスＩＤ・初期標準モデル・参照モデル対応表を用いた初期標準モデルの決定方法を示すフローチャートである。

図６３は、図６２におけるステップＳ１１０５の具体例を示す図である。

図６４は、第３近似部を用いて作成した標準モデルを用いた認識実験の結果を示すグラフである。

図６５（ａ）～（ｊ）は、音声認識の対象についての属性と標準モデルの構造（ガウス分布の混合数）との関係例を示す図である。

発明を実施するための最良の形態

以下、本発明の実施の形態について図面を参照しながら詳しく説明する。なお、図中同一又は相当部分には同一符号を付し、その説明は繰り返さない。

（第１の実施の形態）

図１は、本発明の第１の実施の形態における標準モデル作成装置の全

体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ１０１に組み込まれた例が示されている。本実施の形態では特定の属性を有する音声の特徴を示す音声認識用の標準モデルを作成する場合を例にして説明する。

サーバ１０１は、通信システムにおけるコンピュータ装置等であり、事象の集合と事象又は事象間の遷移の出力確率で表現された隠れマルコフモデルによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、読み込み部１１１と、参照モデル準備部１０２と、参照モデル記憶部１０３と、標準モデル作成部１０４と、書き込み部１１２とを備える。

読み込み部１１１は、ＣＤ－ＲＯＭなどのストレージデバイスに書き込まれた子供用参照モデル、成人用参照モデル、高齢者用参照モデルを読み込む。参照モデル準備部１０２は、読み込まれた参照モデル１２１を参照モデル記憶部１０３へ送信する。参照モデル記憶部１０３は、３個の参照モデル１２１を記憶する。ここで、参照モデルとは、標準モデルを作成するに際して参照される予め作成されたモデル（ここでは、音声認識用のモデル、つまり、一定の属性を有する音声の特徴を示す確率モデル）である。

標準モデル作成部１０４は、参照モデル記憶部１０３が記憶した３個（Ｎｇ＝３）の参照モデル１２１に対する確率又は尤度を最大化又は極大化するように標準モデル１２２を作成する処理部であり、標準モデルの構造（ガウス分布の混合数など）を決定する標準モデル構造決定部１０４ａと、標準モデルを計算するための統計量の初期値を決定することで初期標準モデルを作成する初期標準モデル作成部１０４ｂと、決定された初期標準モデルを記憶する統計量記憶部１０４ｃと、統計量記憶部１０４ｃに記憶された初期標準モデルに対して、第１近似部１０４ｅに

よる近似計算等を用いることにより、参照モデル記憶部１０３に記憶されている３個（Ｎｇ＝３）の参照モデル１２１に対する確率又は尤度を最大化又は極大化するような統計量を算出する（最終的な標準モデルを生成する）統計量推定部１０４ｄとからなる。なお、統計量とは、標準モデルを特定するパラメータであり、ここでは、混合重み係数、平均値、分散値である。

書き込み部１１２は、標準モデル作成部１０４が作成した標準モデル１２２をＣＤ－ＲＯＭなどのストレージデバイスに書き込む。

次に、以上のように構成されたサーバ１０１の動作について説明する。

図２は、サーバ１０１の動作手順を示すフローチャートである。

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（ステップＳ１００）。つまり、読み込み部１１１は、ＣＤ－ＲＯＭなどのストレージデバイスに書き込まれた子供用参照モデル、成人用参照モデル、高齢者用参照モデルを読み込み、参照モデル準備部１０２は、読み込まれた参照モデル１２１を参照モデル記憶部１０３へ送信し、参照モデル記憶部１０３は、３個の参照モデル１２１を記憶する。

参照モデル１２１は、音素ごとのＨＭＭにより構成される。参照モデル１２１の一例を図３に示す。ここでは、子供用参照モデル、成人用参照モデル、高齢者用参照モデルのイメージ図が示されている（なお、本図では、高齢者用参照モデルのイメージ図は省略されている）。これら３個の参照モデルの全てが、状態数３個、各状態は混合分布数が３個の混合ガウス分布によりＨＭＭの出力分布が構成される。特徴量として１２次元（Ｊ＝１２）のケプストラム係数が用いられる。

次に、標準モデル作成部１０４は、参照モデル記憶部１０３が記憶した３個の参照モデル１２１に対する確率又は尤度を最大化又は極大化するように標準モデル１２２を作成する（ステップＳ１０１）。

最後に、書き込み部１１２は、標準モデル作成部１０４が作成した標準モデル１２２をＣＤ－ＲＯＭなどのストレージデバイスに書き込む（ステップＳ１０２）。ＣＤ－ＲＯＭなどのストレージデバイスに書き込まれた標準モデルは、子供、成人、高齢者を考慮した音声認識用の標準モデルとして利用される。

図４は、図２におけるステップＳ１０１（標準モデルの作成）の詳細な手順を示すフローチャートである。

まず、標準モデル構造決定部１０４ａは、標準モデルの構造を決定する（ステップＳ１０２ａ）。ここでは、標準モデルの構造として、音素ごとのＨＭＭにより構成され、３状態であり、各状態における出力分布の混合数を３個（Ｍｆ＝３）と決定する。

次に、初期標準モデル作成部１０４ｂは、標準モデルを計算するための統計量の初期値を決定する（ステップＳ１０２ｂ）。ここでは、参照モデル記憶部１０３に記憶された３つの参照モデルを、統計処理計算を用いて１つのガウス分布に統合したものを統計量の初期値とし、その初期値を初期標準モデルとして統計量記憶部１０４ｃに記憶する。

具体的には、初期標準モデル作成部１０４ｂは、上記３つの状態Ｉ（Ｉ＝１、２、３）それぞれについて、以下の式１３に示される出力分布を生成する。なお、式中のＭｆ（ガウス分布の混合数）は、ここでは、３である。

（式１３）

$$\sum_{m=1}^{M_f} \omega_{f(m)} f(x; \mu_{f(m)}, \sigma^2_{f(m)})$$

ここで、

（式１４）

$$f(x;\mu_{f(m)},\sigma^2_{f(m)}) \quad (m=1,2,...,M_f)$$

は、ガウス分布を表し、

（式１５）

$$x=(x_{(1)},x_{(2)},...,x_{(J)})\in R^J$$

は、１２次元（Ｊ＝１２）のＬＰＣケプストラム係数を表し、

（式１６）

$$\omega_{f(m)} \quad (m=1,2,...,M_f)$$

は、各ガウス分布の混合重み係数を表し、

（式１７）

$$\mu_{f(m)}=(\mu_{f(m,1)},\mu_{f(m,2)},...,\mu_{f(m,J)})\in R^J \quad (m=1,2,...,M_f)$$

は、各ガウス分布の平均値を表し、

（式１８）

$$\sigma^2_{f(m)}=(\sigma^2_{f(m,1)},\sigma^2_{f(m,2)},...,\sigma^2_{f(m,J)})\in R^J \quad (m=1,2,...,M_f)$$

は、各ガウス分布の分散値を表す。

そして、統計量推定部１０４ｄは、参照モデル記憶部１０３に記憶された３つの参照モデル１２１を用いて、統計量記憶部１０４ｃに記憶された標準モデルの統計量を推定する（ステップＳ１０２ｃ）。

具体的には、３つ（Ｎｇ＝３）の参照モデル１２１の各状態Ｉ（Ｉ＝１、２、３）における出力分布、即ち、以下の式１９に示される出力分布に対する標準モデルの確率又は尤度（以下の式２５に示される尤度ｌｏｇＰ）を極大化もしくは最大化するような標準モデルの統計量（上記式

１６に示される混合重み係数、上記式１７に示される平均値、及び、上記式１８に示される分散値）を推定する。

（式１９）

$$\sum_{l=1}^{L_{g(i)}} \upsilon_{g(i,l)} g(x;\mu_{g(i,l)},\sigma^2_{g(i,l)}) \quad (i=1,2,...,N_g)$$

ここで、

（式２０）

$$g(x;\mu_{g(i,l)},\sigma^2_{g(i,l)}) \quad (i=1,2,...,N_g, l=1,2,...,L_{(i)})$$

はガウス分布を表し、

（式２１）

$$L_{g(i)} \quad (i=1,2,...,N_g)$$

は各参照モデルの混合分布数（ここでは、３）を表し、

（式２２）

$$\upsilon_{g(i,l)} \qquad (l=1,2,...,L_{g(i)})$$

は各ガウス分布の混合重み係数を表し、

（式２３）

$$\mu_{g(i,l)} \qquad (l=1,2,...,L_{g(i)})$$

は各ガウス分布の平均値を表し、

（式２４）

$$\sigma^2_{g(i,l)} \qquad (l=1,2,...,L_{g(i)})$$

は各ガウス分布の分散値を表す。

（式２５）

$$\log P = \sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \log[\sum_{m=1}^{M_f} \omega_{f(m)} f(x;\mu_{f(m)},\sigma^2_{f(m)})]\{\sum_{l=1}^{L_{g(i)}} \upsilon_{g(i,l)} g(x;\mu_{g(i,l)},\sigma^2_{g(i,l)})\}dx$$

そして、以下の式２６、式２７及び式２８に従って、それぞれ、標準モデルの混合重み係数、平均値及び分散値を算出する。

（式２６）

$$\omega_{f(m)} = \frac{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \gamma(x,m)\{\sum_{l=1}^{L_{g(i)}} \upsilon_{g(i,l)} g(x;\mu_{g(i,l)},\sigma^2_{g(i,l)})\}dx}{\sum_{k=1}^{M_f}\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \gamma(x,k)\{\sum_{l=1}^{L_{g(i)}} \upsilon_{g(i,l)} g(x;\mu_{g(i,l)},\sigma^2_{g(i,l)})\}dx}$$

$$(m = 1,2,\ldots,M_f)$$

（式２７）

$$\mu_{f(m,j)} = \frac{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \gamma(x,m)x_{(j)}\{\sum_{l=1}^{L_{g(i)}} \upsilon_{g(i,l)} g(x;\mu_{g(i,l)},\sigma^2_{g(i,l)})\}dx}{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \gamma(x,m)\{\sum_{l=1}^{L_{g(i)}} \upsilon_{g(i,l)} g(x;\mu_{g(i,l)},\sigma^2_{g(i,l)})\}dx}$$

$$(m = 1,2,\ldots,M_f, j = 1,2,\ldots,J)$$

（式２８）

$$\sigma^2_{f(m,j)} = \frac{\sum_{l=1}^{N_g} \int_{\infty} \gamma(x,m)(x_{(j)} - \mu_{f(m,j)})^2 \{\sum_{l=1}^{L_{g(l)}} \upsilon_{g(i,l)} g(x;\mu_{g(i,l)}, \sigma^2_{g(i,l)})\} dx}{\sum_{l=1}^{N_g} \int_{\infty} \gamma(x,m) \{\sum_{l=1}^{L_{g(l)}} \upsilon_{g(i,l)} g(x;\mu_{g(i,l)}, \sigma^2_{g(i,l)})\} dx}$$

$$(m = 1,2,\ldots, M_f, j = 1,2,\ldots, J)$$

このとき、統計量推定部１０４ｄの第１近似部１０４ｅにより、以下の式２９に示される近似式が用いられる。

（式２９）

$$\gamma(x,m) = \frac{\omega_{f(m)} f(x;\mu_{f(m)}, \sigma^2_{f(m)})}{\sum_{k=1}^{M_f} \omega_{f(k)} f(x;\mu_{f(k)}, \sigma^2_{f(k)})} \approx \frac{\omega_{f(m)} f(x;\mu_{f(m)}, \sigma^2_{f(m)})}{u_{h(m)} h(x;\mu_{h(m)}, \sigma^2_{h(m)})}$$

$$(m = 1,2,\ldots, M_f)$$

ここで、

（式３０）

$$u_{h(m)} h(x;\mu_{h(m)}, \sigma^2_{h(m)}) \quad (m = 1,2,\ldots, M_f)$$

は、

（式３１）

$$u_{h(m)} \quad (m = 1,2,\ldots, M_f)$$

（式３２）

$$\mu_{h(m)} = (\mu_{h(m,1)}, \mu_{h(m,2)}, \ldots, \mu_{h(m,J)}) \in R^J$$

を平均値とし、

（式３３）

$$\sigma^2_{h(m)} = (\sigma^2_{h(m,1)}, \sigma^2_{h(m,2)}, \cdots, \sigma^2_{h(m,J)}) \in R^J$$

を分散値とする単一のガウス分布を表す。

また、第１近似部１０４ｅは、上記式３０に示された単一ガウス分布の重み（式３１）平均値（式３２）及び分散値（式３３）を、それぞれ、以下の式３４、式３５及び式３６に示された式に従って算出する。

（式３４）

$$u_{h(m)} = \sum_{p=1}^{M_f} \omega_{f(m,p)} = \sum_{p=1}^{M_f} \omega_{f(p)} = 1.0 \quad (m = 1,2,\ldots, M_f)$$

（式３５）

$$\mu_{h(m,j)} = \frac{\sum_{p=1}^{M_f} \omega_{f(m,p)} \mu_{f(m,p,j)}}{\sum_{p=1}^{M_f} \omega_{f(m,p)}} = \frac{\sum_{p=1}^{M_f} \omega_{f(p)} \mu_{f(p,j)}}{\sum_{p=1}^{M_f} \omega_{f(p)}}$$

$$(m = 1,2,\ldots, M_f, j = 1,2,\ldots, J)$$

（式３６）

$$\sigma^2_{h(m,j)} = \frac{\sum_{p=1}^{M_f} \omega_{f(m,p)} \left(\sigma^2_{f(m,p,j)} + \mu^2_{f(m,p,j)} \right)}{\sum_{p=1}^{M_f} \omega_{f(m,p)}} - \mu^2_{h(m,j)}$$

$$= \frac{\sum_{p=1}^{M_f} \omega_{f(p)} \left(\sigma^2_{f(p,j)} + \mu^2_{f(p,j)} \right)}{\sum_{p=1}^{M_f} \omega_{f(p)}} - \mu^2_{h(m,j)}$$

$$(m = 1,2,\ldots, M_f, j = 1,2,\ldots, J)$$

図５は、第１近似部１０４ｅによる近似計算を説明する図である。第１近似部１０４ｅは、本図に示されるように、上記式２９に示された近似式における単一ガウス分布（式３０）を、標準モデルを構成する全ての混合ガウス分布を用いて決定している。

以上の第１近似部１０４ｅによる近似式を考慮してまとめると、統計量推定部１０４ｄでの計算式は次の通りになる。つまり、統計量推定部１０４ｄは、以下の式３７、式３８及び式３９に従って、それぞれ、混合重み係数、平均値及び分散値を算出し、統計量記憶部１０４ｃに記憶する。そして、このような統計量の推定と統計量記憶部１０４ｃへの記憶をＲ（≧１）回、繰り返す。その結果得られた統計量を最終的に生成する標準モデル１２２の統計量として出力する。

（式３７）

$$\omega_{f(m)} = \frac{\sum_{i=1}^{N_g} \prod_{j=1}^{J} \sum_{l=1}^{L_{g(i)}} A_{(m,l,i,j)}}{\sum_{i=1}^{N_g} \sum_{k=1}^{M_f} \prod_{j=1}^{J} \sum_{l=1}^{L_{g(i)}} A_{(k,l,i,j)}} \quad (m = 1,2,\ldots, M_f)$$

$$A_{(m,l,i,j)} = \frac{\omega_{f(m)} \upsilon_{g(i,j)} \sigma^2_{h(m,j)}}{\sqrt{\sigma^2_{f(m,j)}\sigma^2_{h(m,j)} + \sigma^2_{g(i,j,j)}\sigma^2_{h(m,j)} - \sigma^2_{f(m,j)}\sigma^2_{g(i,j,j)}}}$$

$$\times \exp\left\{\frac{1}{2}\left(\frac{\left(\frac{\sigma_{f(m,j)}\sigma_{g(i,j,j)}}{\sigma_{h(m,j)}}\mu_{h(m,j)} - \frac{\sigma_{f(m,j)}\sigma_{h(m,j)}}{\sigma_{g(i,j,j)}}\mu_{g(i_m,j)} - \frac{\sigma_{g(i,j,j)}\sigma_{h(m,j)}}{\sigma_{f(m,j)}}\mu_{f(m,j)}\right)^2}{\sigma^2_{f(m,j)}\sigma^2_{h(m,j)} + \sigma^2_{g(i,j,j)}\sigma^2_{h(m,j)} - \sigma^2_{f(m,j)}\sigma^2_{g(i,j,j)}} + \frac{\mu_{h(m,j)}}{\sigma^2_{h(m,j)}} - \frac{\mu_{g(i_m,j)}}{\sigma^2_{g(i,j,j)}} - \frac{\mu_{f(m,j)}}{\sigma^2_{f(m,j)}}\right)\right\}$$

（式３８）

$$\mu_{f(m,j)} = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{L_{g(i)}} B_{(m,l,i,j)}}{\sum_{i=1}^{N_g} \prod_{j=1}^{J} \sum_{l=1}^{L_{g(i)}} A_{(m,l,i,j)}} \quad (m = 1,2,\ldots, M_f, j = 1,2,\ldots, J)$$

$$B_{(m,l,i,j)} = \frac{\sigma^2_{f(m,j)}\sigma^2_{h(m,j)}\mu_{g(i_m,j)} + \sigma^2_{g(i,l,j)}\sigma^2_{h(m,j)}\mu_{f(m,j)} - \sigma^2_{f(m,j)}\sigma^2_{g(i,l,j)}\mu_{h(m,j)}}{\sigma^2_{f(m,j)}\sigma^2_{h(m,j)} + \sigma^2_{g(i,l,j)}\sigma^2_{h(m,j)} - \sigma^2_{f(m,j)}\sigma^2_{g(i,l,j)}} \times A_{(m,l,i,j)}$$

（式３９）

$$\sigma^2_{f(m,j)} = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{L_{g(i)}} C_{(m,l,i,j)}}{\sum_{i=1}^{N_g} \prod_{j=1}^{J} \sum_{l=1}^{L_{g(i)}} A_{(m,l,i,j)}} \quad (m = 1,2,\ldots, M_f)$$

$$C_{(m,l,i,j)} = \left\{\frac{\sigma^2_{f(m,j)}\sigma^2_{g(i,j,j)}\sigma^2_{h(m,j)}}{\sigma^2_{f(m,j)}\sigma^2_{h(m,j)} + \sigma^2_{g(i,l,j)}\sigma^2_{h(m,j)} - \sigma^2_{f(m,j)}\sigma^2_{g(i,l,j)}}\right.$$

$$\left. + \left(\mu_{f(m,j)} - \frac{\sigma^2_{f(m,j)}\sigma^2_{h(m,j)}\mu_{g(i_m,j)} + \sigma^2_{g(i,j,j)}\sigma^2_{h(m,j)}\mu_{f(m,j)} - \sigma^2_{f(m,j)}\sigma^2_{g(i,l,j)}\mu_{h(m,j)}}{\sigma^2_{f(m,j)}\sigma^2_{h(m,j)} + \sigma^2_{g(i,l,j)}\sigma^2_{h(m,j)} - \sigma^2_{f(m,j)}\sigma^2_{g(i,l,j)}}\right)^2\right\} \times A_{(m,l,i,j)}$$

なお、状態遷移確率については、ＨＭＭの対応する状態遷移確率を参照モデル１２１に対して全て加えあわせた全体が１になるように正規化したものを用いる。

次に、本実施の形態をパーソナルコンピュータによる音声認識に適用した具体例を説明する。ここでは、サーバ１０１としてパソコン（ＰＣ）、読み込み部１１１としてＣＤ－ＲＯＭドライブ装置を用いるものとし、標準モデルの具体的な使い方を中心に説明する。

まず、利用者は、ＰＣ（サーバ１０１）のＣＤ－ＲＯＭドライブ装置（読み込み部１１１）に、参照モデルとしての複数の音響モデルが格納された１枚のＣＤ－ＲＯＭを装着する。そのＣＤ－ＲＯＭには、例えば、「幼児」、「子供：男」、「子供：女」、「大人：男」、「大人：女」、「高齢者：男」、「高齢者：女」の各音響モデルが記憶されている。

次に、利用者は、図６（ａ）及び（ｂ）に示される画面表示例のように、ＰＣ（サーバ１０１）に接続されたディスプレイを用いて、家族構成（音声認識を利用する人）にあった音響モデルを選択する。図６には、ＣＤ－ＲＯＭに記憶されている音響モデルが「ＣＤ－ＲＯＭ」と書かれた枠内に表示され、それらの音響モデルの中から選択された音響モデルが「利用者」と書かれた枠内にコピーされる様子が示されている。ここでは、利用者の家族構成が、１０歳の男の子と、５０歳のお父さんと、４０歳のお母さんの３人であるとし、利用者（お父さん）によって、「子供：男」、「大人：男」、「大人：女」の３個のモデルが「利用者」と書かれた枠内にドラッグして移動されている。このような操作によって、参照モデル準備部１０２による参照モデルの準備が行われる。つまり、３個の参照モデルが読み込み部１１１で読み出され、参照モデル準備部１０２を介して、参照モデル記憶部１０３に格納される。

続いて、利用者は、図７（ａ）に示される画面表示例のように、作成

する標準モデルの構造（混合分布数）を指定する。図７（ａ）では、「混合分布数」として「３個」、「１０個」、「２０個」が表示され、利用者は、これらの個数の中から希望するものを選択する。この操作によって、標準モデル構造決定部１０４ａにより、これから作成する標準モデルの構造が決定される。

なお、混合分布数の決定については、このような直接的な指定に限られず、例えば、図７（ｂ）に示される画面表示例のように、利用者が選択した仕様情報に基づいて混合分布数を決定してもよい。図７（ｂ）では、標準モデルを使用して音声認識を実行させる対象機器として、３種類の「利用機器」、つまり、「テレビ用」、「カーナビ用」、「携帯電話用」の中から利用機器を選択する様子が示されている。このとき、予め記憶された対応表に従って、例えば、「テレビ用」が選択された場合には混合分布数を３個と決定し、「カーナビ用」が選択された場合には混合分布数を２０個と決定し、「携帯電話用」が選択された場合には混合分布数を１０個と決定してもよい。

その他、混合分布数の決定については、認識速度や精度、つまり、「素早く認識」、「通常」、「高精度に認識」の中から選択することで、それぞれの選択項目に対応した値（「素早く認識」＝３個、「通常」＝１０個、「高精度に認識」＝２０個）を混合分布数として決定してもよい。

このような入力操作が終了すると、初期標準モデル作成部１０４ｂによって初期標準モデルが作成された後に、統計量推定部１０４ｄによる繰り返し計算（学習）が行われ、標準モデルが作成される。このとき、図８の画面表示例に示されるように、標準モデル構造決定部１０４ａによって、学習の進捗状況が表示される。利用者は、学習の進捗状況、学習終了時期などを知ることができ、標準モデルが完成されるまで安心して待つことができる。なお、進捗状況の表示として、例えば、図８（ａ）

に示されるような学習度合いのバー表示、図８（ｂ）に示されるような学習回数の表示、その他、尤度基準の表示等がある。また、未学習時は一般的な顔画像を表示し、学習の完了に近づくにしたがって利用者の顔画像に変更していくような進捗表示であってもよい。同様に、未学習時には赤ちゃんを表示し、学習の完了に近づくにしたがって仙人を表示するような進捗表示であってもよい。

このようにして標準モデルの作成が完了すると、作成された標準モデルは、標準モデル作成部１０４によってメモリカード（書き込み部１１２）に記録される。利用者は、そのメモリカードをＰＣ（サーバ１０１の書き込み部１１２）から抜き出し、利用機器、例えば、テレビのメモリカード用スロットに挿入する。これによって、作成された標準モデルがＰＣ（サーバ１０１）から利用機器（テレビ）に移動される。テレビは、装着されたメモリカードに記録された標準モデルを用いて、利用者（ここでは、テレビを利用する家族）を対象とした音声認識を行う。たとえば、テレビに付属したマイクに入力された音声を認識することによって、テレビ操作用のコマンドを判別し、そのコマンド（例えば、チャンネルの切り替え、ＥＰＧなどによる番組検索）を実行する。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを用いた、音声によるテレビ操作が実現される。

以上説明したように、本発明の第１の実施の形態によれば、予め準備された参照モデルに対する確率又は尤度を最大化又は極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、学習のためのデータや教師データを必要とすることなく簡易に標準モデルが作成されるとともに、既に作成された複数の参照モデルを総合的に勘案した精度の高い標準モデルが作成される。

なお、標準モデル１２２は、音素ごとにＨＭＭを構成するに限らず、

文脈依存のＨＭＭで構成してもよい。

また、標準モデル作成部１０４は、一部の音素の、一部の状態における事象の出力確率に対してモデル作成を行ってもよい。

また、標準モデル１２２を構成するＨＭＭは、音素ごとに異なる状態数により構成してもよいし、状態ごとに異なる分布数の混合ガウス分布により構成してもよい。

また、参照モデル１２１は、子供用参照モデル、成人用参照モデル、高齢者用参照モデルにおいて、異なる状態数により構成してもよいし、異なる混合数の混合ガウス分布により構成してもよい。

また、標準モデル１２２を用いて、サーバ１０１において音声認識を行ってもよい。

また、参照モデル１２１をＣＤ－ＲＯＭ、ＤＶＤ－ＲＡＭなどのストレージデバイスから読み込む代わりに、サーバ１０１において音声データから参照モデル１２１を作成してもよい。

また、参照モデル準備部１０２は、必要に応じてＣＤ－ＲＯＭ、ＤＶＤ－ＲＡＭなどのストレージデバイスから読み込まれた新たな参照モデルを参照モデル記憶部１０３に追加・更新してもよい。つまり、参照モデル準備部１０２は、新たな参照モデルを参照モデル記憶部１０３に格納するだけでなく、同一の認識対象についての参照モデルが参照モデル記憶部１０３に格納されている場合には、その参照モデルと置き換えることによって参照モデルを更新したり、参照モデル記憶部１０３に格納されている不要な参照モデルを削除してもよい。

また、参照モデル準備部１０２は、必要に応じて、通信路を介して新たな参照モデルを参照モデル記憶部１０３に追加・更新してもよい。

また、標準モデルを作成したのちに、さらに音声データにより学習してもよい。

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのＨＭＭの構造や、状態数などを決定してもよい。

（第２の実施の形態）

図９は、本発明の第２の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がセットトップボックス２０１（以下、ＳＴＢと呼ぶ）に組み込まれた例が示されている。本実施の形態では音声認識用の標準モデル（話者適応モデル）を作成する場合を例にして説明する。具体的には、ＳＴＢによる音声認識機能により、テレビのＥＰＧ検索や番組切替、録画予約などを行う場合を例にして説明する。

ＳＴＢ２０１は、ユーザの発話を認識してＴＶ番組の自動切替等を行うデジタル放送用受信機であり、事象の集合と事象又は事象間の遷移の出力確率とによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、マイク２１１と、音声データ蓄積部２１２と、参照モデル準備部２０２と、参照モデル記憶部２０３と、利用情報作成部２０４と、参照モデル選択部２０５と、標準モデル作成部２０６と、音声認識部２１３とを備える。

マイク２１１に収集された音声データは、音声データ蓄積部２１２に蓄積される。参照モデル準備部２０２は、音声データ蓄積部２１２が蓄積した音声データを用いて話者ごとに参照モデル２２１を作成し、参照モデル記憶部２０３に記憶する。

利用情報作成部２０４は、利用情報２２４である利用者の音声をマイク２１１により収集する。ここで、利用情報とは、認識（狭義での認識、識別、認証など）の対象（人・物）に関する情報であり、ここでは、音声認識の対象となる利用者の音声である。参照モデル選択部２０５は、利用情報作成部２０４が作成した利用情報２２４に基づいて、参照モデ

ル記憶部２０３が記憶している参照モデル２２１の中から、利用情報２２４が示す利用者の音声に音響的に近い参照モデル２２３を選択する。

標準モデル作成部２０６は、参照モデル選択部２０５が選択した話者の参照モデル２２３に対する確率又は尤度を最大化又は極大化するように標準モデル２２２を作成する処理部であり、標準モデルの構造（ガウス分布の混合分布数など）を決定する標準モデル構造決定部２０６ａと、標準モデルを計算するための統計量の初期値を決定することで初期標準モデルを作成する初期標準モデル作成部２０６ｂと、決定された初期標準モデルを記憶する統計量記憶部２０６ｃと、統計量記憶部２０６ｃに記憶された初期標準モデルに対して、一般近似部２０６ｅによる近似計算等を用いることにより、参照モデル選択部２０５が選択した参照モデル２２３に対する確率又は尤度を最大化又は極大化するような統計量を算出する（最終的な標準モデルを生成する）統計量推定部２０６ｄとからなる。

音声認識部２１３は、標準モデル作成部２０６によって作成された標準モデル２２２を用いて利用者の音声を認識する。

次に、以上のように構成されたＳＴＢ２０１の動作について説明する。

図１０は、ＳＴＢ２０１の動作手順を示すフローチャートである。

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（ステップＳ２００）。つまり、マイク２１１によりＡさんからＺさんの音声データを収集して音声データ蓄積部２１２に蓄積する。たとえば、屋内に設置された複数のマイク、テレビのリモコンに内蔵されたマイク、電話機などが、ＳＴＢ２０１の音声データ蓄積部２１２と接続されており、マイクや電話機から入力された音声データを音声データ蓄積部２１２に蓄積する。たとえば、お兄ちゃん、妹、お父さん、お母さん、おじいちゃん、近所のひと、友達の音声が蓄積される。

参照モデル準備部２０２は、音声データ蓄積部２１２が蓄積した音声データを用いて話者ごとに参照モデル２２１をバウム・ウェルチの再推定の方法により作成する。この処理は、標準モデルの作成が要求される以前に行われる。

参照モデル記憶部２０３は、参照モデル準備部２０２が作成した参照モデル２２１を記憶する。参照モデル２２１は、音素ごとのＨＭＭにより構成される。参照モデル２２１の一例を図１１に示す。ここでは、Ａさんから Ｚさんの全ての参照モデルが、状態数３個、各状態は混合分布数が５個の混合ガウス分布によりＨＭＭの出力分布が構成される。特徴量として２５次元（Ｊ＝２５）のメルケプストラム係数が用いられる。

ここで、標準モデルの作成が要求される。たとえば、利用者が「利用者の確認」のボタンを押すことによって、標準モデルの作成が要求される。「利用者確認」のボタンについては、テレビ画面に表示させて選択する方法や、テレビのリモコンに「利用者の確認」スイッチをつけて選択する方法が考えられる。ボタンを押すタイミングとしては、テレビを起動したタイミング、音声認識を用いてコマンド操作を行っているときに利用者にふさわしい標準モデルがほしいと感じたタイミングなどが考えられる。

次に、利用情報作成部２０４は、利用情報２２４である利用者の音声をマイク２１１により収集する（ステップＳ２０１）。たとえば、標準モデルの作成が要求されると、画面上で「名前を入力してください」と表示される。利用者は、テレビのリモコンに内蔵されたマイクにより名前（利用者の音声）を入力する。この利用者の音声が利用情報である。なお、入力する音声は名前に限定されない。例えば「適応と発声してください」と表示して、利用者は「適応」と発声してもよい。

参照モデル選択部２０５は、その利用者の音声に音響的に近い参照モ

デル２２３を、参照モデル記憶部２０３が記憶している参照モデル２２１の中から選択する（ステップＳ２０２）。具体的には、利用者の音声をＡさんからＺさんの参照モデルに入力して発声単語に対する尤度が大きい１０人（Ｎｇ＝１０）の話者の参照モデルを選択する。

そして、標準モデル作成部２０６は、参照モデル選択部２０５が選択した１０個の参照モデル２２３に対する確率又は尤度を最大化又は極大化するように標準モデル２２２を作成する（ステップＳ２０３）。このとき、第１の実施の形態のように、学習の進捗状況を表示してもよい。そうすることで、利用者は学習の進捗状況、学習終了時期などが判断でき、安心して標準モデルを作成することができる。また、学習の進捗状況を非表示にする進捗状況非表示部を設けてもよい。この機能により、画面を有効に使うことができる。また、慣れた人に対して非表示にすることで、うっとうしく感じることが回避される。

最後に、音声認識部２１３は、マイク２１１から介して送られてくる利用者の音声を入力とし、標準モデル作成部２０６で作成された標準モデル２２２を用いて音声認識を行う（Ｓ２０４）。たとえば、利用者が発話した音声を音響解析等を行うことで２５次元のメルケプストラム係数を算出し、音素ごとの標準モデル２２２に入力することで、高い尤度を有する音素の連なりを特定する。そして、その音素の連なりと予め受信している電子番組データ中の番組名とを比較し、一定以上の尤度が検出された場合に、その番組に切り替えるという自動番組切替の制御を行う。

次に、図１０におけるステップＳ２０３（標準モデルの作成）の詳細な手順を説明する。手順の流れは、図４に示されたフローチャートと同様である。ただし、採用する標準モデルの構造や具体的な近似計算等が異なる。

まず、標準モデル構造決定部２０６ａは、標準モデルの構造を決定す

る（図４のステップＳ１０２ａ）。ここでは、標準モデルの構造として、音素ごとのＨＭＭにより構成され、３状態であり、各状態における出力分布の混合分布数が１６個（Ｍｆ＝１６）と決定する。

次に、初期標準モデル作成部２０６ｂは、標準モデルを計算するための統計量の初期値を決定する（図４のステップＳ１０２ｂ）。ここでは、参照モデル選択部２０５が選択した１０個の参照モデル２２３を、統計処理計算を用いて１つのガウス分布に統合したものを統計量の初期値とし、その初期値を初期標準モデルとして統計量記憶部２０６ｃに記憶する。ここでは、話者ごとに学習した混合分布数が５の参照モデルを用いて精度の高い混合分布数が１６（１６混合）の標準モデル（話者適応モデル）を作成する。

具体的には、初期標準モデル作成部２０６ｂは、上記３つの状態Ｉ（Ｉ＝１、２、３）それぞれについて、上記式１３に示される出力分布を生成する。

ただし、本実施の形態では、上記式１３に示された出力分布における（式４０）

$$x = (x_{(1)}, x_{(2)}, \cdots, x_{(J)}) \in R^J$$

は、２５次元（Ｊ＝２５）のメルケプストラム係数を表す。

そして、統計量推定部２０６ｄは、参照モデル選択部２０５が選択した１０個の参照モデル２２３を用いて、統計量記憶部２０６ｃに記憶された標準モデルの統計量を推定する（図４のステップＳ１０２ｃ）。

つまり、１０個（Ｎｇ＝１０）の参照モデル２２３の各状態Ｉ（Ｉ＝１、２、３）における出力分布、即ち、上記式１９に示される出力分布に対する標準モデルの確率（ここでは、上記式２５に示される尤度 log Ｐ）を極大化もしくは最大化するような標準モデルの統計量（上記式１

６に示される混合重み係数、上記式１７に示される平均値、及び、上記式１８に示される分散値）を推定する。

ただし、本実施の形態では、上記式１９に示された出力分布における

（式４１）

$$L_{g(i)} \quad (i=1,2,...,N_g)$$

は、５（各参照モデルの混合分布数）である。

具体的には、上記式２６、式２７及び式２８に従って、それぞれ、標準モデルの混合重み係数、平均値及び分散値を算出する。

このとき、統計量推定部２０６ｄの一般近似部２０６ｅにより、上記式２９に示される近似式が用いられる。

ここで、一般近似部２０６ｅは、第１の実施の形態と異なり、上記式２９の近似式の分母に示された出力分布

（式４２）

$$\omega_{f(k)} f(x;\mu_{f(k)},\sigma^2_{f(k)}) \quad (k=1,2,...,M_f)$$

の中から、上記式２９の近似式の分子に示された出力分布

（式４３）

$$\omega_{f(m)} f(x;\mu_{f(m)},\sigma^2_{f(m)})$$

に距離的に近い３個（Ｐｈ(m)＝３）の出力分布

（式４４）

$$\omega_{f(m,p)} f(x;\mu_{f(m,p)},\sigma^2_{f(m,p)}) \quad (m=1,2,...,M_f, p=1,2,...,P_{h(m)})$$

を選択し、選択した３個の出力分布を用いて、上記式３０に示された単一ガウス分布の重み（式３１）、平均値（式３２）及び分散値（式３３）

を、それぞれ、以下の式４５、式４６及び式４７に示された式に従って算出する。

（式４５）

$$u_{h(m)} = \sum_{p=1}^{P_{h(m)}} \omega_{f(m,p)} \quad (m = 1,2,\ldots, M_f)$$

（式４６）

$$\mu_{h(m,J)} = \frac{\sum_{p=1}^{P_{h(m)}} \omega_{f(m,p)} \mu_{f(m,p,j)}}{\sum_{p=1}^{P_{h(m)}} \omega_{f(m,p)}} \quad (m = 1,2,\ldots, M_f, j = 1,2,\ldots, J)$$

（式４７）

$$\sigma^2_{h(m,j)} = \frac{\sum_{p=1}^{P_{h(m)}} \omega_{f(m,p)} \left(\sigma^2_{f(m,p,j)} + \mu^2_{f(m,p,j)}\right)}{\sum_{p=1}^{P_{h(m)}} \omega_{f(m,p)}} - \mu^2_{h(m,j)}$$

$$(m = 1,2,\ldots, M_f, j = 1,2,\ldots, J)$$

図１２は、一般近似部２０６ｅによる近似計算を説明する図である。一般近似部２０６ｅは、本図に示されるように、上記式２９に示された近似式における単一ガウス分布（式３０）を、標準モデルを構成するＭｆ個の混合ガウス分布の中から、計算対象となる混合ガウス分布に近い一部（Ｐｈ(m)個）の混合ガウス分布だけを用いて決定している。したがって、全部（Ｍｆ個）の混合ガウス分布を用いる第１の実施の形態と比較し、近似計算における計算量が削減される。

以上の一般近似部２０６ｅによる近似式を考慮してまとめると、統計量推定部２０６ｄでの計算式は次の通りになる。つまり、統計量推定部

２０６ｄは、以下の式４８、式４９及び式５０に従って、それぞれ、混合重み係数、平均値及び分散値を算出し、統計量記憶部２０６ｃに記憶する。そして、このような統計量の推定と統計量記憶部２０６ｃへの記憶をＲ（≧１）回、繰り返す。その結果得られた統計量を最終的に生成する標準モデル２２２の統計量として出力する。なお、繰り返し計算においては、その回数に対応させて、上記近似計算における出力分布の選択個数Ｐｈ(m)を小さくし、最終的にＰｈ(m)＝１とする計算を行う。

（式４８）

$$\omega_{f(m)} = \frac{\sum_{l=1}^{N_g} \sum_{l=1}^{L_{g(l)}} \alpha_{(m,l,i)}}{\sum_{k=1}^{M_f} \omega_{f(k)} \left(\sum_{l=1}^{N_g} \sum_{l=1}^{L_{g(l)}} \alpha_{(k,l,i)} \right)} \quad (m = 1,2,\ldots, M_f)$$

$$\alpha_{(m,l,i)} = \upsilon_{g(i,l)} \prod_{j=1}^{J} D_{(m,l,i,j)}$$

$$D_{(m,l,j)} = \frac{\sigma^2_{h(m,j)}}{\sqrt{\sigma^2_{f(m,j)}\sigma^2_{h(m,j)} + \sigma^2_{g(i,l,j)}\sigma^2_{h(m,j)} - \sigma^2_{f(m,j)}\sigma^2_{g(i,l,j)}}}$$

$$\times \exp\left\{ \frac{1}{2} \left(\frac{\left(\frac{\sigma_{f(m,j)}\sigma_{g(i,l,j)}}{\sigma_{h(m,j)}} \mu_{h(m,j)} - \frac{\sigma_{f(m,j)}\sigma_{h(m,j)}}{\sigma_{g(i,l,j)}} \mu_{g(i,m,j)} - \frac{\sigma_{g(i,l,j)}\sigma_{h(m,j)}}{\sigma_{f(m,j)}} \mu_{f(m,j)} \right)^2}{\sigma^2_{f(m,j)}\sigma^2_{h(m,j)} + \sigma^2_{g(i,l,j)}\sigma^2_{h(m,j)} - \sigma^2_{f(m,j)}\sigma^2_{g(i,l,j)}} + \frac{\mu_{h(m,j)}}{\sigma^2_{h(m,j)}} \frac{\mu_{g(i,m,j)}}{\sigma^2_{g(i,l,j)}} - \frac{\mu_{f(m,j)}}{\sigma^2_{f(m,j)}} \right) \right\}$$

（式４９）

$$\mu_{f(m,j)} = \frac{\sum_{i=1}^{N_g}\sum_{l=1}^{L_{g(i)}} \beta_{(m,l,i,j)}\alpha_{(m,l,i)}}{\sum_{i=1}^{N_g}\sum_{l=1}^{L_{g(i)}} \alpha_{(m,l,i)}} \quad (m=1,2,...,M_f, j=1,2,...,J)$$

$$\beta_{(m,l,i,j)} = \frac{\sigma^2_{f(m,j)}\sigma^2_{h(m,j)}\mu_{g(i,m,j)} + \sigma^2_{g(i,l,j)}\sigma^2_{h(m,j)}\mu_{f(m,j)} - \sigma^2_{f(m,j)}\sigma^2_{g(i,l,j)}\mu_{h(m,j)}}{\sigma^2_{f(m,j)}\sigma^2_{h(m,j)} + \sigma^2_{g(i,l,j)}\sigma^2_{h(m,j)} - \sigma^2_{f(m,j)}\sigma^2_{g(i,l,j)}}$$

（式５０）

$$\sigma^2_{f(m,j)} = \frac{\sum_{i=1}^{N_g}\sum_{l=1}^{L_{g(i)}} \gamma_{(m,l,i,j)}\alpha_{(m,l,i)}}{\sum_{i=1}^{N_g}\sum_{l=1}^{L_{g(i)}} \alpha_{(m,l,i)}} \quad (m=1,2,...,M_f, j=1,2,...,J)$$

$$\gamma_{(m,l,i,j)} = \{\frac{\sigma^2_{f(m,j)}\sigma^2_{g(i,l,j)}\sigma^2_{h(m,j)}}{\sigma^2_{f(m,j)}\sigma^2_{h(m,j)} + \sigma^2_{g(i,l,j)}\sigma^2_{h(m,j)} - \sigma^2_{f(m,j)}\sigma^2_{g(i,l,j)}}$$

$$+\left(\mu_{f(m,j)} - \frac{\sigma^2_{f(m,j)}\sigma^2_{h(m,j)}\mu_{g(i,m,j)} + \sigma^2_{g(i,l,j)}\sigma^2_{h(m,j)}\mu_{f(m,j)} - \sigma^2_{f(m,j)}\sigma^2_{g(i,l,j)}\mu_{h(m,j)}}{\sigma^2_{f(m,j)}\sigma^2_{h(m,j)} + \sigma^2_{g(i,l,j)}\sigma^2_{h(m,j)} - \sigma^2_{f(m,j)}\sigma^2_{g(i,l,j)}}\right)^2\}$$

なお、状態遷移確率については、ＨＭＭの対応する状態遷移確率を参照モデル２２３に対して全て加えあわせた全体が１になるように正規化したものを用いる。

以上説明したように、本発明の第２の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率又は尤度を最大化又は極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、利用状況によりふさわしい精度の高い標準モデルが提供される。

なお、標準モデルを作成するタイミングとしては、本実施の形態のよ

うな利用者による明示的な指示だけに限られず、他のタイミングで標準モデルを作成してもよい。たとえば、ＳＴＢ２０１にさらに、利用者が変更されたかどうかを自動的に判断する利用者変更判断部を設ける。その利用者変更判断部は、テレビのリモコンに入力された認識用の音声を用いて、利用者が変更されたか否か、つまり、現在の利用者が直前まで認識していた利用者と同一人物であるか否かを判断する。利用者が変更されたと判断した場合に、その音声を利用情報として標準モデルを作成する。これにより、利用者が意識することなく、利用者にふさわしい標準モデルを用いた音声認識が行われる。

なお、標準モデル２２２は、音素ごとにＨＭＭを構成するに限らず、文脈依存のＨＭＭで構成してもよい。

また、標準モデル作成部２０６は、一部の音素の、一部の状態における事象の出力確率に対してモデル作成を行ってもよい。

また、標準モデル２２２を構成するＨＭＭは、音素ごとに異なる状態数により構成してもよいし、状態ごとに異なる分布数の混合ガウス分布により構成してもよい。

また、参照モデル２２１は、話者ごとＨＭＭにおいて、異なる状態数により構成してもよいし、異なる混合数の混合ガウス分布により構成してもよい。

また、参照モデル２２１は、話者ごとＨＭＭに限らず、話者・雑音・声の調子ごとに作成してもよい。

また、標準モデル２２２をＣＤ－ＲＯＭ、ハードディスク、ＤＶＤ－ＲＡＭなどのストレージデバイスに記録してもよい。

また、参照モデル２２１を作成する代わりに、ＣＤ－ＲＯＭ、ＤＶＤ－ＲＡＭなどのストレージデバイスから読み込んでもよい。

また、参照モデル選択部２０５は、利用情報２２４に基づいて利用者

ごとに選択する参照モデルの数を変えてもよい。

また、参照モデル準備部２０２は、必要に応じて新たな参照モデルを作成して参照モデル記憶部２０３に追加・更新してもよいし、参照モデル記憶部２０３に格納されている不要な参照モデルを削除してもよい。

また、参照モデル準備部２０２は、必要に応じて、通信路を介して新たな参照モデルを参照モデル記憶部２０３に追加・更新してもよい。

また、上記近似計算において選択する出力分布の個数Ｐh(m)は、対象とする事象や標準モデルの出力分布によって異なってもよいし、分布間距離に基づいて決定してもよい。

また、標準モデルを作成したのちに、さらに音声データにより学習してもよい。

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのＨＭＭの構造や、状態数などを決定してもよい。

また、混合分布数については、本実施の形態におけるＳＴＢを出荷するときに、所定の値に設定しておいてもよいし、ネットワーク連携を考慮した機器のＣＰＵパワーなどの仕様、起動するアプリケーションの仕様などに基づいて混合分布数を決定してもよい。

（第３の実施の形態）

図１３は、本発明の第３の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がＰＤＡ（Ｐｅｒｓｏｎａｌ　Ｄｉｇｉｔａｌ　Ａｓｓｉｓｔａｎｔ）３０１に組み込まれた例が示されている。本実施の形態では雑音識別用の標準モデル（雑音モデル）を作成する場合を例にして説明する。

ＰＤＡ３０１は、携帯情報端末であり、事象の出力確率によって定義される雑音識別用の標準モデルを作成する標準モデル作成装置として、

読み込み部３１１と、参照モデル準備部３０２と、参照モデル記憶部３０３と、利用情報作成部３０４と、参照モデル選択部３０５と、標準モデル作成部３０６と、仕様情報作成部３０７と、マイク３１２と、雑音識別部３１３とを備える。

読み込み部３１１は、ＣＤ－ＲＯＭなどのストレージデバイスに書き込まれた乗用車Ａの参照モデル、乗用車Ｂの参照モデル、バスＡの参照モデル、小雨の参照モデル、大雨の参照モデルなどの雑音の参照モデルを読み込む。参照モデル準備部３０２は、読み込まれた参照モデル３２１を参照モデル記憶部３０３へ送信する。参照モデル記憶部３０３は、参照モデル３２１を記憶する。

利用情報作成部３０４は、利用情報３２４である雑音の種類をＰＤＡ３０１の画面とキーを利用して作成する。参照モデル選択部３０５は、利用情報３２４である雑音の種類に音響的に近い参照モデルを、参照モデル記憶部３０３が記憶している参照モデル３２１の中から選択する。仕様情報作成部３０７は、ＰＤＡ３０１の仕様に基づき仕様情報３２５を作成する。ここで、仕様情報とは、作成する標準モデルの仕様に関する情報であり、ここでは、ＰＤＡ３０１が備えるＣＰＵの処理能力に関する情報である。

標準モデル作成部３０６は、仕様情報作成部３０７で作成された仕様情報３２５に基づいて、参照モデル選択部３０５が選択した雑音の参照モデル３２３に対する確率又は尤度を最大化又は極大化するように標準モデル３２２を作成する処理部であり、標準モデルの構造（ガウス分布の混合分布数など）を決定する標準モデル構造決定部３０６ａと、標準モデルを計算するための統計量の初期値を決定することで初期標準モデルを作成する初期標準モデル作成部３０６ｂと、決定された初期標準モ

された初期標準モデルに対して、第２近似部３０６ｅによる近似計算等を用いることにより、参照モデル選択部３０５が選択した参照モデル３２３に対する確率又は尤度を最大化又は極大化するような統計量を算出する（最終的な標準モデルを生成する）統計量推定部３０６ｄとからなる。

雑音識別部３１３は、標準モデル作成部３０６で作成された標準モデル３２２を用いて、マイク３１２から入力された雑音の種類を識別する。

次に、以上のように構成されたＰＤＡ３０１の動作について説明する。

図１４は、ＰＤＡ３０１の動作手順を示すフローチャートである。

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（ステップＳ３００）。つまり、読み込み部３１１は、ストレージデバイスに書き込まれた雑音の参照モデルを読み込み、参照モデル準備部３０２は、読み込まれた参照モデル３２１を参照モデル記憶部３０３へ送信し、参照モデル記憶部３０３は、参照モデル３２１を記憶する。

参照モデル３２１は、ＧＭＭより構成される。参照モデル３２１の一例を図１５に示す。ここでは、各雑音モデルは混合分布数が３個のＧＭＭにより構成される。特徴量として５次元（Ｊ＝５）のＬＰＣケプストラム係数が用いられる。

次に、利用情報作成部３０４は、識別したい雑音の種類である利用情報３２４を作成する（ステップＳ３０１）。図１６にＰＤＡ３０１の選択画面の一例を示す。ここでは、乗用車の雑音が選択される。参照モデル選択部３０５は、選択された利用情報３２４である乗用車の雑音に音響的に近い参照モデルである乗用車Ａの参照モデルと乗用車Ｂの参照モデルを、参照モデル記憶部３０３が記憶している参照モデル３２１の中から選択する（ステップＳ３０２）。

様情報３２５を作成する（ステップＳ３０３）。ここでは、ＰＤＡ３０１のＣＰＵの仕様に基づきＣＰＵパワーが小さいという仕様情報３２５を作成する。標準モデル作成部３０６は、作成された仕様情報３２５に基づいて、参照モデル選択部３０５が選択した参照モデル３２３に対する確率又は尤度を最大化又は極大化するように標準モデル３２２を作成する（ステップＳ３０４）。

最後に、雑音識別部３１３は、利用者によってマイク３１２から入力された雑音に対して、標準モデル３２２を用いて、雑音の識別を行う（ステップＳ３０５）。

次に、図１４におけるステップＳ３０４（標準モデルの作成）の詳細な手順を説明する。手順の流れは、図４に示されたフローチャートと同様である。ただし、採用する標準モデルの構造や具体的な近似計算等が異なる。

まず、標準モデル構造決定部３０６ａは、標準モデルの構造を決定する（図４のステップＳ１０２ａ）。ここでは、標準モデルの構造として、仕様情報３２５であるＣＰＵパワーが小さいという情報に基づいて１混合（Ｍｆ＝１）のＧＭＭにより標準モデル３２２を構成すると決定する。

次に、初期標準モデル作成部３０６ｂは、標準モデルを計算するための統計量の初期値を決定する（図４のステップＳ１０２ｂ）。ここでは、選択された参照モデル３２３である乗用車Ａの３混合の参照モデルを、統計処理計算を用いて１つのガウス分布に統合したものを統計量の初期値として統計量記憶部３０６ｃに記憶する。

具体的には、初期標準モデル作成部３０６ｂは、上記式１３に示される出力分布を生成する。

ただし、本実施の形態では、上記式１３に示された出力分布における

（式５１）

$$x = (x_{(1)}, x_{(2)}, \ldots, x_{(J)}) \in R^J$$

は、５次元（Ｊ＝５）のＬＰＣケプストラム係数を表す。

そして、統計量推定部３０６ｄは、参照モデル選択部３０５が選択した２個の参照モデル３２３を用いて、統計量記憶部３０６ｃに記憶された標準モデルの統計量を推定する（図４のステップＳ１０２ｃ）。

つまり、２個（Ｎｇ＝２）の参照モデル３２３における出力分布、即ち、上記式１９に示される出力分布に対する標準モデルの確率（ここでは、上記式２５に示される尤度 logＰ）を極大化もしくは最大化するような標準モデルの統計量（上記式１６に示される混合重み係数、上記式１７に示される平均値、及び、上記式１８に示される分散値）を推定する。

ただし、本実施の形態では、上記式１９に示された出力分布における

（式５２）

$$L_{g(i)} \quad (i = 1,2,\ldots,N_g)$$

は、３（各参照モデルの混合分布数）である。

具体的には、上記式２６、式２７及び式２８に従って、それぞれ、標準モデルの混合重み係数、平均値及び分散値を算出する。

このとき、統計量推定部３０６ｄの第２近似部３０６ｅは、標準モデルの各ガウス分布はお互いに影響を与えないと仮定して、以下の近似式を用いる。

（式５３）

$$\gamma(x,m) \approx \frac{\omega_{f(m)} f(x;\mu_{f(m)},\sigma^2_{f(m)})}{u_{h(m)} h(x;\mu_{h(m)},\sigma^2_{h(m)})} \approx 1.0$$

$$(m = 1,2,\ldots,M_f)$$

また、標準モデルのガウス分布

（式５４）

$$\omega_{f(m,p)} f(x;\mu_{f(m,p)},\sigma^2_{f(m,p)}) \quad (m = 1,2,\ldots,M_f, p = 1,2,\ldots,P_{h(m)})$$

の近傍の

（式５５）

$$x$$

とは、前記式５４が示す出力分布との平均値のユークリッド距離、マハラノビス距離、カルバック・ライブラー（ＫＬ）距離などの分布間距離が近いＱg(m,i)個の参照モデル３２３のガウス分布

（式５６）

$$g(x;\mu_{g(i,l)},\sigma^2_{g(i,l)}) \quad (i = 1,2,\ldots,N_g, l = 1,2,\ldots,L_{(i)})$$

が存在する空間であって、

（式５７）

$$\omega_{f(m,p)} f(x;\mu_{f(m,p)},\sigma^2_{f(m,p)}) \quad (m = 1,2,\ldots,M_f, p = 1,2,\ldots,P_{h(m)})$$

との分布間距離が近いＱg(m,i)個（１≦Ｑg(m,i)≦Ｌg(i)）の前記参照ベクトルの出力分布とは、前記参照モデルの出力分布

（式５８）

$$\upsilon_{g(i,l)} g(x;\mu_{g(l)},\sigma^2_{g(l)}) \quad (i=1,2,...,N_g, l=1,2,...,L_{g(i)})$$

のうち分布間距離が１番近い（近傍指示パラメータＧ＝１）前記標準モデルの出力分布が前記式５７である前記参照ベクトルの出力分布であると近似する。

図１７は、この統計量推定部３０６ｄによる統計量の推定手順を示す概念図である。各参照モデルの各ガウス分布に対して、平均値のユークリッド距離、マハラノビス距離などの分布間距離が最も近いものが標準モデルのガウス分布 m であるガウス分布を用いて統計量の推定を行うことが示されている。

図１８は、第２近似部３０６ｅによる近似計算を説明する図である。第２近似部３０６ｅは、本図に示されるように、各参照モデルの各ガウス分布に対して、距離が最も近い標準モデルのガウス分布ｍを決定することで、上記式５３に示された近似式を用いている。

以上の第２近似部３０６ｅによる近似式を考慮してまとめると、統計量推定部３０６ｄでの計算式は次の通りになる。つまり、統計量推定部３０６ｄは、以下の式５９、式６０及び式６１に従って、それぞれ、混合重み係数、平均値及び分散値を算出し、それらのパラメータによって特定される標準モデルを最終的な標準モデル３２２として生成する。

（式５９）

$$\omega_{f(m)} = \frac{\sum_{i=1}^{N_g}\sum_{l=1}^{Q_{g(m,l)}}\upsilon_{g(i,l)}}{\sum_{k=1}^{M_f}\sum_{i=1}^{N_g}\sum_{l=1}^{Q_{g(m,l)}}\upsilon_{g(i,l)}}$$

$$(m=1,2,...,M_f)$$

（ここで、分母、分子の和は、各参照モデルの各ガウス分布に対して、平均値のユークリッド距離、マハラノビス距離などの分布間距離が最も近いものが標準モデルのガウス分布 m であるガウス分布に関する和を意味する。）

（式６０）

$$\mu_{f(m,j)} = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{Q_{g(m,i)}} \upsilon_{g(i,l)} \mu_{g(i,l,j)}}{\sum_{i=1}^{N_g} \sum_{l=1}^{Q_{g(m,l)}} \upsilon_{g(i,l)}}$$

$$(m = 1,2,\ldots,M_f, j = 1,2,\ldots,J)$$

（ここで、分母、分子の和は、各参照モデルの各ガウス分布に対して、平均値のユークリッド距離、マハラノビス距離などの分布間距離が最も近いものが標準モデルのガウス分布 m であるガウス分布に関する和を意味する。）

（式６１）

$$\sigma^2_{f(m,j)} = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{Q_{g(m,j)}} \upsilon_{g(i,l)} \left(\sigma^2_{g(i,l)} + \mu^2_{g(i,l,j)} \right)}{\sum_{i=1}^{N_g} \sum_{l=1}^{Q_{g(m,i)}} \upsilon_{g(i,l)}} - \mu^2_{f(m,j)}$$

$$(m = 1,2,\ldots,M_f, j = 1,2,\ldots,J)$$

（ここで、分母、分子の和は、各参照モデルの各ガウス分布に対して、平均値のユークリッド距離、マハラノビス距離などの分布間距離が最も近いものが標準モデルのガウス分布 m であるガウス分布に関する和を

ただし、

（式６２）

$$\sum_{i=1}^{N_g} Q_{g(m,i)} = 0 \quad (m = 1,2,\ldots, M_f)$$

の場合において、

（第１の方法）混合重み係数、平均値、分散値を更新しない。

（第２の方法）混合重み係数の値をゼロにして、平均値、分散値を所定の値にする。

（第３の方法）混合重み係数の値を所定の値にして、平均値、分散値を標準モデルの出力分布を１個の分布に表現したときの平均値、分散値にする。

のいずれかを利用して統計量の値を決定する。なお、利用する方法は、繰り返し回数Ｒ、ＨＭＭ、ＨＭＭの状態ごとに異なっていてもよい。ここでは、第１の方法を用いる。

統計量推定部３０６ｄは、このように推定した標準モデルの統計量を統計量記憶部３０６ｃに記憶する。そして、このような統計量の推定と統計量記憶部３０６ｃへの記憶をＲ（≧１）回、繰り返す。その結果得られた統計量を最終的に生成する標準モデル３２２の統計量として出力する。

次に、本実施の形態をＰＤＡによる環境音識別に適用した具体例を説明する。

まず、参照モデル準備部３０２は、ＣＤ－ＲＯＭから環境音の識別に必要な参照モデルを読み出す。利用者は、識別を行う環境（利用情報）を考慮して、識別したい環境音を画面上から選択する。たとえば、「乗用車」を選択し、続いて、「警報音」、「赤ちゃんの声」、「電車の音」などを

選択する。この選択に基づいて、参照モデル選択部３０５は、参照モデル記憶部３０３に記憶されている参照モデルの中から対応する参照モデルを選択する。そして、選択した参照モデル３２３を１つずつ用いて、標準モデル作成部３０６は、それぞれに対して標準モデルを作成する。

続いて、利用者は、ＰＤＡ３０１において、「らくらく情報提供」（環境音に基づく状況判断による情報提供）というアプリケーションプログラムを起動する。このアプリケーションは、環境音に基づいて状況判断を行い、利用者に適切な情報を提供するプログラムである。起動されると、ＰＤＡ３０１の表示画面に「正確に判断」、「素早く判断」という表示がされる。これに対して、利用者はどちらかを選択する。

そして、仕様情報作成部３０７は、その選択結果に基づいて、仕様情報を作成する。たとえば、「正確に判断」が選択された場合には、精度を高くするために、混合分布数を１０個とする仕様情報を作成する。一方、「素早く判断」が選択された場合には、高速に処理するために、混合分布数を１個とする仕様情報を作成する。なお、複数のＰＤＡが連携して処理できる場合などには、現在利用できるＣＰＵパワーを判断し、そのＣＰＵパワーに基づいて仕様情報を作成してもよい。

このような仕様情報にしたがって、「乗用車」、「警報音」、「赤ちゃんの声」、「電車の音」などの１混合の標準モデルが作成される。そして、ＰＤＡ３０１は、作成された標準モデルにより環境識別を行い、その識別結果に基づき、各種情報をＰＤＡの画面に表示する。例えば、「乗用車」が近くにあると識別した場合は、道路地図を表示したり、「赤ちゃんの声」を識別した場合は、おもちゃ屋さんの広告を表示したりする。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを用いた、環境音識別に基づく情報提供が実現される。なお、アプリケーションの仕様に応じて標準モデルの複雑さを調節することが

できる。

以上説明したように、本発明の第３の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率又は尤度を最大化又は極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、利用状況によりふさわしい精度の高い標準モデルが提供される。

また、仕様情報に基づいて標準モデルが作成されるため、標準モデルを利用する機器にふさわしい標準モデルが準備される。

なお、統計量推定部３０６ｄによる処理の繰り返し回数は、上記式２５に示された尤度の大きさがある一定のしきい値以上になるまでの回数としてもよい。

また、標準モデル３２２を構成するＧＭＭは、雑音の種類ごとに異なる混合分布数の混合ガウス分布により構成してもよい。

また、識別モデルは、雑音モデルに限らず、話者を識別してもよいし、年齢などを識別してもよい。

また、標準モデル３２２をＣＤ－ＲＯＭ、ＤＶＤ－ＲＡＭ、ハードディスクなどのストレージデバイスに記録してもよい。

また、参照モデル３２１をＣＤ－ＲＯＭなどのストレージデバイスから読み込む代わりに、ＰＤＡ３０１において雑音データから参照モデル３２１を作成してもよい。

また、参照モデル準備部３０２は、必要に応じてＣＤ－ＲＯＭなどのストレージデバイスから読み込まれた新たな参照モデルを参照モデル記憶部３０３に追加・更新してもよいし、参照モデル記憶部３０３に格納されている不要な参照モデルを削除してもよい。

また、参照モデル準備部３０２は、必要に応じて、通信路を介して新たな参照モデルを参照モデル記憶部３０３に追加・更新してもよい。

また、標準モデルを作成したのちに、さらにデータにより学習してもよい。

また、標準モデル構造決定部は、標準モデルの構造や、状態数などを決定してもよい。

また、近傍指示パラメータＧは、対象とする事象や標準モデルの出力分布によって異なってもよいし、繰り返し回数Ｒによって変化させてもよい。

（第４の実施の形態）

図１９は、本発明の第４の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ４０１に組み込まれた例が示されている。本実施の形態では顔認識用の標準モデルを作成する場合を例にして説明する。

サーバ４０１は、通信システムにおけるコンピュータ装置等であり、事象の出力確率によって定義される顔認識用の標準モデルを作成する標準モデル作成装置として、カメラ４１１と、画像データ蓄積部４１２と、参照モデル準備部４０２と、参照モデル記憶部４０３と、利用情報受信部４０４と、参照モデル選択部４０５と、標準モデル作成部４０６と、書き込み部４１３とを備える。

カメラ４１１により、顔の画像データが収集され、画像データ蓄積部４１２に顔画像データが蓄積される。参照モデル準備部４０２は、画像データ蓄積部４１２が蓄積した顔画像データを用いて話者ごとに参照モデル４２１を作成し、参照モデル記憶部４０３に記憶する。

利用情報受信部４０４は、利用者が希望する顔認識の対象となる人間の年齢の年代と性別の情報を利用情報４２４として電話４１４により受信する。参照モデル選択部４０５は、利用情報受信部４０４が受信した

利用情報４２４に基づいて、参照モデル記憶部４０３が記憶している参照モデル４２１の中から、利用情報４２４が示す年代と性別の話者に対応する参照モデル４２３を選択する。

標準モデル作成部４０６は、参照モデル選択部４０５が選択した話者の顔画像の参照モデル４２３に対する確率又は尤度を最大化又は極大化するように標準モデル４２２を作成する処理部であり、第２の実施の形態における標準モデル作成部２０６と同一の機能を有するとともに、第１の実施の形態における第１近似部１０４ｅと第３の実施の形態における第２近似部３０６ｅの機能を有する。つまり、第１～第３の実施の形態で示された３種類の近似計算を組み合わせた計算を行う。

書き込み部４１３は、標準モデル作成部４０６が作成した標準モデル４２２をＣＤ－ＲＯＭなどのストレージデバイスに書き込む。

次に、以上のように構成されたサーバ４０１の動作について説明する。

図２０は、サーバ４０１の動作手順を示すフローチャートである。図２１は、サーバ４０１の動作手順を説明するための参照モデル及び標準モデルの一例を示す図である。

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（図２０のステップＳ４００）。つまり、カメラ４１１によりＡさんからＺさんの顔画像データを収集して画像データ蓄積部４１２に蓄積する。参照モデル準備部４０２は、画像データ蓄積部４１２が蓄積した顔画像データを用いて、話者ごとの参照モデル４２１をＥＭアルゴリズムにより作成する。ここでは参照モデル４２１はＧＭＭで構成される。

参照モデル記憶部４０３は、参照モデル準備部４０２が作成した参照モデル４２１を記憶する。ここでは、図２１の参照モデル４２１に示されるように、ＡさんからＺさんの全ての参照モデルが、混合分布数が５個のＧＭＭにより構成される。特徴量として１００次元（Ｊ＝１００）

の画素の濃度値を用いる。

次に、利用情報受信部４０４は、利用情報４２４である年代と性別の情報を電話４１４により受信する（図２０のステップＳ４０１）。ここでは、利用情報４２４として、１１歳から１５歳の男性と２２歳から２６歳の女性である。参照モデル選択部４０５は、その利用情報４２４に基づいて、参照モデル記憶部４０３が記憶している参照モデル４２１から、利用情報４２４に対応する参照モデル４２３を選択する（図２０のステップＳ４０２）。具体的には、図２１の「選択された参照モデル４２３」に示されるように、ここでは、１１歳から１５歳の男性及び２２歳から２６歳の女性の参照モデルを選択する。

そして、標準モデル作成部４０６は、参照モデル選択部４０５が選択した話者の参照モデル４２３に対する確率又は尤度を最大化又は極大化するように標準モデル４２２を作成する（図２０のステップＳ４０３）。ここでは、図２１の標準モデル４２２に示されるように、２つの標準モデル４２２それぞれを、混合分布数が３個のＧＭＭにより構成する。

標準モデル４２２の作成方法は、基本的には、第２の実施の形態と同様に行われる。ただし、標準モデル４２２の統計量の推定における近似計算については、具体体には、以下のようにして行われる。つまり、標準モデル作成部４０６は、内蔵の記憶部等を介することで、第１の実施の形態における第１近似部１０４ｅによる近似計算と同様の近似計算よって作成したモデルを初期値として、第２の実施の形態における一般近似部２０６ｅによる近似計算と同様の近似計算による計算を行い、その結果を初期値として第３の実施の形態における第２近似部３０６ｅによる近似計算と同様の近似計算を行う。

書き込み部４１３は、標準モデル作成部４０６が作成した２つの標準モデル４２２をＣＤ－ＲＯＭなどのストレージデバイスに書き込む（図

２０のステップＳ４０４）。

利用者は、１１歳から１５歳の男性の標準モデルと２２歳から２６歳の女性の標準モデルが書き込まれたストレージデバイスを郵送で受け取る。

次に、本実施の形態を、行動予測に基づいてお店などを紹介する情報提供システムに適用した具体例を説明する。この情報提供システムは、通信ネットワークで接続されたカーナビゲーション装置と情報提供サーバ装置から構成される。カーナビゲーション装置は、本実施の形態における標準モデル作成装置４０１によって予め作成された標準モデルを行動予測モデルとして利用することで、人の行動（つまり、車による行先等）を予測し、その行動に関連した情報（行先の近くに位置するレストランなどのお店の情報など）を提供する機能を備える。

まず、利用者は、カーナビゲーション装置を用いて、電話回線４１４で接続されたサーバ４０１に対して、自分用の行動予測モデルの作成を依頼する。

具体的には、利用者は、カーナビゲーション装置が表示する項目選択画面で、「らくらく推薦機能」のボタンを押す。すると、利用者の住所（利用場所）、年齢、性別、趣味などを入力する画面になる。

ここでは、利用者はお父さんとお母さんとする。まず、お父さんの個人情報をカーナビゲーション装置の画面と対話しながら入力する。住所については、電話番号を入力することにより自動的に変換される。あるいは、カーナビゲーション装置において現在位置が表示されているときに「利用場所」のボタンを押すことで、その現在位置が利用場所として入力される。ここでは住所の情報を住所Ａとする。年齢と性別については、「５０代」、「男」を選択して入力する。趣味については、予め表示されたチェック項目があるので、利用者は、該当箇所をチェックする。こ

こではお父さんの趣味の情報を趣味情報Ａとする。

続いて、お母さんの個人情報についても同様に入力する。住所Ｂ、４０代、女、趣味情報Ｂからなる個人情報が作成される。このような入力の結果は、図２２の画面表示例に示されるとおりである。

最後に、カーナビゲーション装置は、このようにして作成された個人情報を利用情報として、付属の電話回線４１４を用いて、情報提供サーバ装置であるサーバ４０１に転送する。

次に、サーバ４０１は、転送されてきた個人情報（利用情報）に基づいて、お父さんとお母さんの２個の行動予測モデルを作成する。ここで、行動予測モデルは、確率モデルで表現され、その入力は、曜日、時刻、現在地などで、出力は、お店Ａの情報を提示する確率、お店Ｂの情報を提示する確率、お店Ｃの情報を提示する確率、駐車場の情報を提示する確率などである。

サーバ４０１の参照モデル記憶部４０３に記憶されている複数の参照モデルは、年代、性別、代表的な住所と趣味の傾向で作成した行動予測モデルである。サーバ４０１では、予め、カメラ４１１に代えて、カーナビゲーション装置の入力ボタン等を用いて各種個人情報（上記入力及び出力についての情報）を入力することで、画像データ蓄積部４１２に各種個人情報を蓄積したうえで、参照モデル準備部４０２によって、画像データ蓄積部４１２に蓄積された個人情報から、複数種類の典型的な利用者ごとの参照モデル４２１が作成され、参照モデル記憶部４０３に格納されている。

参照モデル選択部４０５は、個人情報（利用情報）を用いて、個人情報にふさわしい参照モデルを選択する。例えば、同じ町の、年代と性別が同じで、趣味のチェック項目が８割以上一致した参照モデルを選択する。サーバ４０１の標準モデル作成部４０６は、選択された参照モデル

を統合した標準モデルを作成する。作成された標準モデルは書き込み部４１３により、メモリカードに記憶される。ここでは、お父さんとお母さんの２人の標準モデルが記憶される。メモリカードは、郵送で利用者に届けられる。

利用者は、受け取ったメモリカードをカーナビゲーション装置に挿入し、画面に表示された「お父さん」と「お母さん」を選択することで、利用者を設定する。これによって、カーナビゲーション装置は、装着されたメモリカードに記憶された標準モデルを行動予測モデルとして使用することで、現在の曜日、時刻、場所などから、必要なタイミングでお店の情報などを提示する。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを行動予測モデルとして用いることで、人の行動（つまり、車による行先）を予測し、その行動に関連した情報を提供する情報提供システムが実現される。

以上説明したように、本発明の第４の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率又は尤度を最大化又は極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、利用状況によりふさわしい高精度な標準モデルが提供される。

なお、標準モデル４２２を構成するＧＭＭは、話者ごとに異なる分布数の混合ガウス分布により構成してもよい。

また、参照モデル準備部４０２は、必要に応じて新たな参照モデルを作成して参照モデル記憶部４０３に追加・更新してもよいし、参照モデル記憶部４０３に格納されている不要な参照モデルを削除してもよい。

また、標準モデルを作成したのちに、さらにデータにより学習してもよい。

また、標準モデル構造決定部は、標準モデルの構造や、状態数などを

決定してもよい。

（第５の実施の形態）

図２３は、本発明の第５の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ５０１に組み込まれた例が示されている。本実施の形態では音声認識用の標準モデル（適応モデル）を作成する場合を例にして説明する。

サーバ５０１は、通信システムにおけるコンピュータ装置等であり、事象の集合と事象又は事象間の遷移の出力確率とによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、読み込み部５１１と、音声データ蓄積部５１２と、参照モデル準備部５０２と、参照モデル記憶部５０３と、利用情報受信部５０４と、参照モデル選択部５０５と、標準モデル作成部５０６と、仕様情報受信部５０７と、書き込み部５１３とを備える。

読み込み部５１１は、ＣＤ－ＲＯＭなどのストレージデバイスに書き込まれた子供、成人、高齢者の音声データを読み込み、音声データ蓄積部５１２に蓄積する。参照モデル準備部５０２は、音声データ蓄積部５１２が蓄積した音声データを用いて話者ごとに参照モデル５２１を作成する。参照モデル記憶部５０３は、参照モデル準備部５０２が作成した参照モデル５２１を記憶する。

仕様情報受信部５０７は、仕様情報５２５を受信する。利用情報受信部５０４は、利用情報５２４である利用者の音声を受信する。参照モデル選択部５０５は、利用情報５２４である利用者の音声に音響的に近い話者の参照モデルを、参照モデル記憶部５０３が記憶している参照モデル５２１から選択する。

標準モデル作成部５０６は、仕様情報５２５に基づいて、参照モデル

選択部５０５が選択した話者の参照モデル５２３に対する確率又は尤度を最大化又は極大化するように標準モデル５２２を作成する処理部であり、第１の実施の形態における標準モデル作成部１０４と同一の機能を有する。書き込み部５１３は、標準モデル作成部５０６が作成した標準モデル５２２をＣＤ－ＲＯＭなどのストレージデバイスに書き込む。

次に、以上のように構成されたサーバ５０１の動作について説明する。

図２４は、サーバ５０１の動作手順を示すフローチャートである。図２５は、サーバ５０１の動作手順を説明するための参照モデル及び標準モデルの一例を示す図である。

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（図２４のステップＳ５００）。つまり、読み込み部５１１は、ＣＤ－ＲＯＭなどのストレージデバイスに書き込まれた音声データを読み込み、音声データ蓄積部５１２に蓄積する。参照モデル準備部５０２は、音声データ蓄積部５１２が蓄積した音声データを用いて話者ごとに参照モデル５２１をバウム・ウェルチの再推定の方法により作成する。参照モデル記憶部５０３は、参照モデル準備部５０２が作成した参照モデル５２１を記憶する。

参照モデル５２１は、音素ごとのＨＭＭにより構成される。ここでは、図２５の参照モデル５２１に示されるように、子供の各話者の参照モデルは、状態数３個、各状態は混合分布数が３個の混合ガウス分布によりＨＭＭの出力分布が構成され、成人の各話者の参照モデルが、状態数３個、各状態は混合分布数が６４個の混合ガウス分布によりＨＭＭの出力分布が構成され、高齢者の各話者の参照モデルは、状態数３個、各状態は混合分布数が１６個の混合ガウス分布によりＨＭＭの出力分布が構成される。これは、子供の音声データが比較的少なく、成人の音声データが多いためである。特徴量として２５次元（Ｊ＝２５）のメルケプスト

ラム係数が用いられる。

次に、利用情報受信部５０４は、利用者の音声を、端末装置５１４から、利用情報５２４として受信する（図２４のステップＳ５０１）。参照モデル選択部５０５は、利用情報５２４である利用者の音声に音響的に近い参照モデル５２３を、参照モデル記憶部５０３が記憶している参照モデル５２１から選択する（図２４のステップＳ５０２）。具体的には、図２５の「選択された参照モデル５２３」に示されるように、ここでは、近い話者１０人（Ｎｇ＝１０）の参照モデルが選択される

そして、仕様情報受信部５０７は、利用者の要求に基づき仕様情報５２５を端末装置５１４から受信する（図２４のステップＳ５０３）。ここでは、速い認識処理という仕様情報５２５を受信する。標準モデル作成部５０６は、仕様情報受信部５０７が受信した仕様情報５２５に基づいて、参照モデル選択部５０５が選択した話者の参照モデル５２３に対する確率又は尤度を最大化又は極大化するように標準モデル５２２を作成する（図２４のステップＳ５０４）。具体的には、標準モデル５２２は、図２５の標準モデル５２２に示されるように、仕様情報５２５である速い認識処理という情報に基づいて、２混合（Ｍｆ＝２）で、３状態のＨＭＭより構成する。ＨＭＭは音素ごとに構成する。

標準モデル５２２の作成方法は、第１の実施の形態と同様に行われる。

書き込み部５１３は、標準モデル作成部５０６が作成した標準モデル５２２をＣＤ－ＲＯＭなどのストレージデバイスに書き込む（図２４のステップＳ５０５）。

次に、本実施の形態を、通信ネットワークを用いた音声認識によるゲームに適用した具体例を説明する。ここでは、サーバ５０１は、作成した標準モデルを用いて音声認識を行う音声認識部を備えるものとする。また、端末装置５１４として、ＰＤＡとする。これらは、通信ネットワ

ークで接続されている。

サーバ５０１では、読み込み部５１１、音声データ蓄積部５１２及び参照モデル準備部５０２により、音声データをＣＤやＤＶＤなどで入手したタイミングで参照モデルを逐次準備している。

利用者は、ＰＤＡ（端末装置５１４）において、音声認識を利用したゲームプログラム、ここでは、「アクションゲーム」を立ち上げる。すると、「『アクション』と発声してください」と表示されるので、利用者は、「アクション」と発声する。その音声は、利用情報として、ＰＤＡ（端末装置５１４）からサーバ５０１に送信され、サーバ５０１の利用情報受信部５０４及び参照モデル選択部５０５により、参照モデル記憶部５０３に記憶された複数の参照モデルの中から利用者に合った参照モデルを選択する。

また、利用者は、速くリアクションしてほしいので、ＰＤＡ（端末装置５１４）の設定画面において「高速に認識する」と設定する。その設定内容は、仕様情報として、ＰＤＡ（端末装置５１４）からサーバ５０１に送信され、サーバ５０１においては、このような仕様情報及び選択された参照モデルに基づいて、標準モデル作成部５０６により、２混合の標準モデルが作成される。

利用者は、アクションゲームにおいて、ＰＤＡのマイクに「右に移動」、「左に移動」などのコマンドを発声する。入力された音声は、サーバへ送信され、既に作成された標準モデルを利用した音声認識が行われる。その認識結果は、サーバ５０１からＰＤＡ（端末装置５１４）に送信され、ＰＤＡ（端末装置５１４）において、送信されてきた認識結果に基づいて、アクションゲームのキャラクタが動く。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを音声認識に用いることで、音声によるアクションゲームが実現される。

また、同様にして、本実施の形態を別のアプリケーション、例えば、通信ネットワークを用いた翻訳システムに適用することもできる。たとえば、利用者は、ＰＤＡ（端末装置５１４）において、「音声翻訳」というアプリケーションプログラムを立ち上げる。すると、「『翻訳』と発声してください」と表示される。利用者は、「翻訳」と発声する。その音声は、利用情報として、ＰＤＡ（端末装置５１４）からサーバ５０１に送信される。また、利用者は、正確に認識してほしいので、そのアプリケーションにおいて、「正確に認識してほしい」旨を指示する。その指示は、仕様情報として、ＰＤＡ（端末装置５１４）からサーバ５０１に送信される。サーバ５０１では、送信されてきた利用情報及び仕様情報に従って、たとえば、１００混合の標準モデルが作成される。

利用者は、ＰＤＡ（端末装置５１４）のマイクに向かって「おはようございます」と発声する。入力された音声はＰＤＡ（端末装置５１４）からサーバ５０１に送信され、サーバ５０１で「おはようございます」と認識された後に、その認識結果がＰＤＡ（端末装置５１４）に返信される。ＰＤＡ（端末装置５１４）は、サーバ５０１から受信した認識結果を英語に翻訳し、その結果「ＧＯＯＤ　ＭＯＲＮＩＮＧ」を画面に表示する。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを音声認識に用いることで、音声による翻訳装置が実現される。

以上説明したように、本発明の第５の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率又は尤度を最大化又は極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、利用状況によりふさわしい精度の高い標準モデルが提供される。

また、仕様情報に基づいて標準モデルが作成されるため、標準モデル

を利用する機器にふさわしい標準モデルが準備される。

また、参照モデル準備部５０２において、参照モデルごとにデータ数に適した混合分布数の精度の高い参照モデルを準備でき、精度の高い参照モデルを用いて標準モデルを作成できる。このため精度の高い標準モデルの利用が可能となる。

なお、標準モデル５２２は、音素ごとにＨＭＭを構成するに限らず、文脈依存のＨＭＭで構成してもよい。

また、標準モデル５２２を構成するＨＭＭは、状態ごとに異なる分布数の混合ガウス分布により構成してもよい。

また、標準モデル５２２を用いて、サーバ５０１において音声認識を行ってもよい。

また、参照モデル準備部５０２は、必要に応じて新たな参照モデルを作成して参照モデル記憶部５０３に追加・更新してもよいし、参照モデル記憶部５０３に格納されている不要な参照モデルを削除してもよい。

また、標準モデルを作成したのちに、さらにデータにより学習してもよい。

また、標準モデル構造決定部は、標準モデルの構造や、状態数などを決定してもよい。

（第６の実施の形態）

図２６は、本発明の第６の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ６０１に組み込まれた例が示されている。本実施の形態では意図理解のための標準モデル（嗜好モデル）を作成する場合を例にして説明する。

サーバ６０１は、通信システムにおけるコンピュータ装置等であり、事象の出力確率によって定義される意図理解用の標準モデルを作成する

標準モデル作成装置として、読み込み部６１１と、参照モデル準備部６０２と、参照モデル記憶部６０３と、利用情報受信部６０４と、参照モデル選択部６０５と、標準モデル作成部６０６と、仕様情報作成部６０７とを備える。

読み込み部６１１は、ＣＤ－ＲＯＭなどのストレージデバイスに書き込まれた年齢別の話者Ａさんから話者Ｚさんの嗜好モデルを読み込み、参照モデル準備部６０２は、読み込まれた参照モデル６２１を参照モデル記憶部６０３へ送信し、参照モデル記憶部６０３は、参照モデル６２１を記憶する。

仕様情報作成部６０７は、普及しているコンピュータのＣＰＵパワーに合わせて仕様情報６２５を作成する。利用情報受信部６０４は、端末装置６１４から利用情報６２４を受信する。参照モデル選択部６０５は、利用情報受信部６０４が受信した利用情報６２４に基づいて、参照モデル記憶部６０３が記憶している参照モデル６２１からの中から、利用情報６２４に対応した参照モデル６２３を選択する。

標準モデル作成部６０６は、仕様情報作成部６０７が作成した仕様情報６２５に基づいて、参照モデル選択部６０５が選択した参照モデル６２３に対する確率又は尤度を最大化又は極大化するように標準モデル６２２を作成する処理部であり、第２の実施の形態における標準モデル作成部２０６と同一の機能を有するとともに、第３の実施の形態における第２近似部３０６ｅの機能を有する。つまり、第２及び第３の実施の形態で示された２種類の近似計算を組み合わせた計算を行う。

次に、以上のように構成されたサーバ６０１の動作について説明する。

図２７は、サーバ６０１の動作手順を示すフローチャートである。図２８は、サーバ６０１の動作手順を説明するための参照モデル及び標準モデルの一例を示す図である。

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（図２７のステップＳ６００）。つまり、読み込み部６１１は、ＣＤ－ＲＯＭなどのストレージデバイスに書き込まれた年齢別の話者Ａさんから話者Ｚさんの嗜好モデルを読み込み、参照モデル準備部６０２は、読み込まれた参照モデル６２１を参照モデル記憶部６０３へ送信し、参照モデル記憶部６０３は、参照モデル６２１を記憶する。

参照モデル６２１は、ＧＭＭより構成される。ここでは、図２８の参照モデル６２１に示されるように、混合分布数が３個のＧＭＭにより構成される。学習データとして、趣味、性格などを数値化した５次元（Ｊ＝５）の特徴量を用いる。参照モデルの準備は、標準モデルの作成が要求される以前に行う。

次に、利用情報受信部６０４は、嗜好モデルを作成したい年齢層である利用情報６２４を受信する（図２７のステップＳ６０１）。ここでは、２０代、３０代、４０代の年代別の嗜好モデルを利用するという利用情報６２４である。参照モデル選択部６０５は、図２８の「選択された参照モデル６２３」に示されるように、利用情報受信部６０４が受信した利用情報６２４が示す年代の話者の嗜好モデルを、参照モデル記憶部６０３が記憶している参照モデル６２１から選択する（図２７のステップＳ６０２）。

そして、仕様情報作成部６０７は、普及しているコンピュータのＣＰＵパワー、記憶容量などに基づき仕様情報６２５を作成する（図２７のステップＳ６０３）。ここでは、通常速度の認識処理という仕様情報６２５を作成する。

標準モデル作成部６０６は、仕様情報作成部６０７が作成した仕様情報６２５に基づいて、参照モデル選択部６０５が選択した話者の参照モデル６２３に対する確率又は尤度を最大化又は極大化するように標準モ

デル６２２を作成する（図２７のステップＳ６０４）。ここでは、標準モデル６２２は、図２８の標準モデル６２２に示されるように、仕様情報６２５である通常速度の認識処理という情報に基づいて３混合（Ｍｆ＝３）のＧＭＭより構成する。

標準モデル６２２の作成方法は、基本的には、第２の実施の形態と同様に行われる。ただし、標準モデル６２２の統計量の推定における近似計算については、具体体には、以下のようにして行われる。つまり、標準モデル作成部６０６は、内蔵の記憶部等を介することで、第２の実施の形態における一般近似部２０６ｅによる近似計算と同様の近似計算による計算を行い、その結果を初期値として第３の実施の形態における第２近似部３０６ｅによる近似計算と同様の近似計算を行う。

次に、本実施の形態を情報検索装置に適用した具体例を説明する。ここでは、参照モデルは、入力が検索キーワードであり、出力が検索ルールＡ、検索ルールＢなどを利用する確率である。異なる検索ルールを用いると、表示される検索結果が異なってくる。また、サーバ６０１の参照モデル記憶部６０３に準備される参照モデルは、代表的な特徴をもつ話者のモデルとする。

まず、利用者は、サーバ６０１に付属しているリモコン（端末装置６１４）を用いて利用情報を入力する。利用情報は、年齢、性格、性別、趣味などでである。また、「子供」、「俳優」、「高校生」などの所定のグループを識別する情報であってもよい。

続いて、利用者は、選択画面で、「カーナビゲーション装置用」、「携帯電話用」、「パソコン用」、「テレビ用」などから１つの利用機器を選択する。サーバ６０１の仕様情報作成部６０７は、利用機器のＣＰＵパワー、記憶容量に基づいて仕様情報を作成する。ここでは、「テレビ用」が選択されたとし、ＣＰＵパワーと記憶容量が小さい旨の仕様情報６２５が作

成され、その仕様情報６２５に基づいて、標準モデル作成部６０６によって、小さいＣＰＵパワーでも動作する３混合の標準モデルが作成される。作成された標準モデルはメモリカードに書き込まれ、そのメモリカードは利用者によってテレビに挿入される。

利用者は、テレビに表示されたＥＰＧなどで、おすすめ番組を検索するために検索キーワードを入力する。すると、テレビは、メモリカードに記録された標準モデルを用いて、検索キーワードに合った検索ルールを決定し、その検索ルールに沿って番組を検索し、利用者の嗜好にあった番組として表示する。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを用いた便利な検索装置が実現される。

以上説明したように、本発明の第６の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率又は尤度を最大化又は極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、利用状況によりふさわしい精度の高い標準モデルが提供される。

また、仕様情報に基づいて標準モデルが作成されるため、標準モデルを利用する機器にふさわしい標準モデルが準備される。

なお、標準モデル６２２を構成するＧＭＭは、話者ごとに異なる分布数の混合ガウス分布により構成してもよい。

また、参照モデル準備部６０２は、必要に応じてＣＤ－ＲＯＭなどのストレージデバイスから読み込まれた新たな参照モデルを参照モデル記憶部６０３に追加・更新してもよいし、参照モデル記憶部６０３に格納されている不要な参照モデルを削除してもよい。

また、参照モデル及び標準モデルのＧＭＭはベイジアンネットの一部を表現するものでもよい。

また、標準モデルを作成したのちに、さらにデータにより学習してもよい。

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのＨＭＭの構造や、状態数などを決定してもよい。

（第７の実施の形態）

図２９は、本発明の第７の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ７０１に組み込まれた例が示されている。本実施の形態では音声認識用の標準モデル（適応モデル）を作成する場合を例にして説明する。

サーバ７０１は、通信システムにおけるコンピュータ装置等であり、事象の集合と事象又は事象間の遷移の出力確率とによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、読み込み部７１１と、参照モデル準備部７０２と、参照モデル記憶部７０３と、利用情報受信部７０４と、参照モデル選択部７０５と、標準モデル作成部７０６と、仕様情報受信部７０７と、標準モデル記憶部７０８と、標準モデル送信部７０９とを備える。

参照モデル準備部７０２は、読み込み部７１１が読み込んだ、ＣＤ－ＲＯＭなどのストレージデバイスに書き込まれた話者・雑音・声の調子別の音声認識用参照モデルを参照モデル記憶部７０３へ送信し、参照モデル記憶部７０３は、送信された参照モデル７２１を記憶する。

仕様情報受信部７０７は、端末装置７１２から仕様情報７２５を受信する。利用情報受信部７０４は、端末装置７１２から利用情報７２４である雑音下で発声した利用者の音声を受信する。参照モデル選択部７０５は、利用情報７２４である利用者の音声に音響的に近い話者・雑音・声調子の参照モデル７２３を、参照モデル記憶部７０３が記憶している

参照モデル７２１の中から選択する。

標準モデル作成部７０６は、仕様情報受信部７０７が受信した仕様情報７２５に基づいて、参照モデル選択部７０５が選択した参照モデル７２３に対する確率又は尤度を最大化又は極大化するように標準モデル７２２を作成する処理部であり、第２の実施の形態における標準モデル作成部２０６と同一の機能を有する。標準モデル記憶部７０８は、仕様情報７２５に基づいた１もしくは複数の標準モデルを記憶する。標準モデル送信部７０９は、利用者の端末装置７１２から仕様情報と標準モデルの要求信号を受信すると、その仕様情報に適した標準モデルを端末装置７１２へ送信する。

次に、以上のように構成されたサーバ７０１の動作について説明する。

図３０は、サーバ７０１の動作手順を示すフローチャートである。図３１は、サーバ７０１の動作手順を説明するための参照モデル及び標準モデルの一例を示す図である。

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する(図３０のステップＳ７００)。つまり、参照モデル準備部７０２は、読み込み部７１１が読み込んだ、ＣＤ－ＲＯＭなどのストレージデバイスに書き込まれた話者・雑音・声の調子別の音声認識用参照モデルを参照モデル記憶部７０３へ送信し、参照モデル記憶部７０３は、送信された参照モデル７２１を記憶する。ここでは、参照モデル７２１は、話者・雑音・声の調子ごとに、音素ごとのＨＭＭにより構成される。また、各参照モデルは、図３１の参照モデル７２１に示されるように、状態数３個、各状態は混合分布数が１２８個の混合ガウス分布によりＨＭＭの出力分布が構成される。特徴量として２５次元（Ｊ＝２５）のメルケプストラム係数が用いられる。

次に、利用情報受信部７０４は、利用者Ａの雑音下での音声を端末装

置７１２から利用情報７２４として受信する（図３０のステップＳ７０１）。参照モデル選択部７０５は、利用情報７２４である利用者Ａの音声に音響的に近い参照モデル７２３を、参照モデル記憶部７０３が記憶している参照モデル７２１の中から選択する(図３０のステップＳ７０２)。具体的には、図３１の「選択された参照モデル７２３」に示されるように、ここでは、近い話者１００人（Ｎｇ＝１００）の参照モデルが選択される

そして、仕様情報受信部７０７は、利用者Ａの要求に基づき仕様情報７２５を端末装置７１２から受信する(図３０のステップＳ７０３)。ここでは、高い認識精度という仕様情報７２５を受信する。標準モデル作成部７０６は、仕様情報７２５に基づいて、参照モデル選択部７０５が選択した参照モデル７２３に対する確率又は尤度を最大化又は極大化するように標準モデル７２２を作成する(図３０のステップＳ７０４)。具体的には、標準モデル７２２は、図３１の標準モデル７２２に示されるように、仕様情報７２５である高い認識精度という情報に基づいて、６４混合（Ｍｆ＝６４）で、３状態のＨＭＭより構成する。ＨＭＭは音素ごとに構成する。

標準モデル７２２の作成方法は、第２の実施の形態と同様に行われる。

標準モデル記憶部７０８は、仕様情報７２５に基づいた１もしくは複数の標準モデル７２２を記憶する。ここでは、以前に作成した標準モデルである利用者Ｂの１６混合のＨＭＭがすでに記憶されており、新たに利用者Ａの６４混合のＨＭＭが記憶される。

利用者Ａは、端末装置７１２からサーバ７０１の標準モデル送信部７０９へ、仕様情報である利用者Ａと雑音の種類と標準モデルの要求信号とを送信する(図３０のステップＳ７０６)。標準モデル送信部７０９は、

その端末装置７１２へ、仕様に適した標準モデルを端末装置７１２へ送信する（図３０のステップＳ７０７）。ここでは、先ほど作成した利用者Ａの標準モデル７２２を端末装置７１２へ送信する。

利用者Ａは端末装置７１２において受信した標準モデル７２２を用いて音声認識を行う（図３０のステップＳ７０８）。

次に、本実施の形態を、通信ネットワークで接続されたカーナビゲーション装置（端末装置７１２）とサーバ装置（サーバ７０１；標準モデル作成装置）から構成される音声認識システムに適用した具体例を説明する。

まず、利用者は、カーナビゲーション装置（端末装置７１２）の画面にて「自分の音声モデルを獲得」する旨のボタンを選択する。すると、「名前を入力」と表示されるので、ボタン操作により自分の名前を入力する。次に、「『音声』と発声してください」と表示されるので、利用者は、カーナビゲーション装置付属のマイクに向かって「音声」と発声する。これらの情報（利用者の名前、雑音下での音声）は、利用情報として、カーナビゲーション装置（端末装置７１２）からサーバ７０１に送信される。

同様にして、利用者は、カーナビゲーション装置（端末装置７１２）の画面にて「高精度の音声認識」のボタンを選択する。すると、その選択情報は、仕様情報として、カーナビゲーション装置（端末装置７１２）からサーバ７０１に送信される。

サーバ７０１は、それらの利用情報及び仕様情報に基づいて、利用者にふさわしい音声認識用の標準モデルを作成し、作成した標準モデルを利用者の名前と対応づけて標準モデル記憶部７０８に格納しておく。

次回にカーナビゲーション装置（端末装置７１２）を起動すると、「名

の名前がサーバ７０１に送信され、標準モデル７２２に格納された対応する標準モデルが標準モデル送信部７０９によってサーバ７０１から端末装置７１２に送信される。名前（利用者）に対応した標準モデルをサーバ７０１からダウンロードした端末装置７１２は、その標準モデルを用いて、利用者に対する音声認識を行い、音声による目的地設定などを行う。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを音声認識に用いることで、音声によってカーナビゲーション装置を操作することが可能となる。

以上説明したように、本発明の第７の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率又は尤度を最大化又は極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、利用状況によりふさわしい精度の高い標準モデルが提供される。

また、仕様情報に基づいて標準モデルが作成されるため、標準モデルを利用する機器にふさわしい標準モデルが準備される。

また、標準モデル記憶部７０８は、複数の標準モデルを記憶することができるため、必要に応じてすぐに標準モデルが提供される。

また、標準モデル送信部７０９により、標準モデルが端末装置７１２へ送信されるので、端末装置７１２とサーバ７０１が空間的に離れた場所に設置してある場合に、端末装置７１２は、容易にサーバ７０１が作成した標準モデルを利用することできる。

なお、標準モデル７２２は、音素ごとにＨＭＭを構成するに限らず、文脈依存のＨＭＭで構成してもよい。

また、標準モデル７２２を構成するＨＭＭは、状態ごとに異なる混合数の混合ガウス分布により構成してもよい。

また、標準モデル７２２を用いて、サーバ７０１において音声認識を

行い、認識結果を端末装置７１２へ送信してもよい。

また、参照モデル準備部７０２は、必要に応じて新たな参照モデルを作成して参照モデル記憶部７０３に追加・更新してもよいし、参照モデル記憶部７０３に格納されている不要な参照モデルを削除してもよい。

また、参照モデル準備部７０２は、必要に応じて、通信路を介して新たな参照モデルを参照モデル記憶部７０３に追加・更新してもよい。

また、標準モデルを作成したのちに、さらにデータにより学習してもよい。

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのＨＭＭの構造や、状態数などを決定してもよい。

（第８の実施の形態）

図３２は、本発明の第８の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置が携帯電話９０１に組み込まれた例が示されている。本実施の形態では音声認識用の標準モデルを作成する場合を例にして説明する。

携帯電話９０１は、携帯情報端末であり、事象の集合と事象又は事象間の遷移の出力確率で表現された隠れマルコフモデルによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、参照モデル受信部９０９と、参照モデル準備部９０２と、参照モデル記憶部９０３と、利用情報作成部９０４と、参照モデル選択部９０５と、類似度情報作成部９０８と、標準モデル作成部９０６と、仕様情報作成部９０７と、マイク９１２と、音声認識部９１３とを備える。

利用情報作成部９０４は、利用情報９２４を携帯電話９０１の画面とキーを利用して作成する。

仕様情報作成部９０７は、携帯電話９０１の仕様に基づき仕様情報９２５を作成する。ここで、仕様情報とは、作成する標準モデルの仕様に

関する情報であり、ここでは、携帯電話９０１が備えるＣＰＵの処理能力に関する情報である。

類似度情報作成部９０８は、利用情報９２４と仕様情報９２５と参照モデル記憶部９０３が記憶した参照モデル９２１に基づいて、類似度情報９２６を作成して参照モデル準備部に送信する。

参照モデル準備部９０２は、類似度情報９２６に基づいて、参照モデルを準備するか否かを決定する。参照モデル準備部９０２は、参照モデルを準備すると決定した場合に、利用情報９２４と仕様情報９２５を参照モデル受信部９０９に送信する。

参照モデル受信部９０９は、利用情報９２４と仕様情報９２５に対応した参照モデルを、サーバ装置９１０から受信して参照モデル準備部９０２に送信する。

参照モデル準備部９０２は、参照モデル受信部９０９が送信した参照モデルを参照モデル記憶部９０３に記憶する。

参照モデル選択部９０５は、利用情報９２４に対応した参照モデル９２３を、参照モデル記憶部９０３が記憶している参照モデル９２１の中から選択する。

標準モデル作成部９０６は、仕様情報作成部９０７で作成された仕様情報９２５に基づいて、参照モデル選択部９０５が選択した参照モデル９２３に対する確率又は尤度を最大化又は極大化するように標準モデル９２２を作成する処理部であり、標準モデルの構造（ガウス分布の混合分布数など）を決定する標準モデル構造決定部９０６ａと、標準モデルを計算するための統計量の初期値を決定することで初期標準モデルを作成する初期標準モデル作成部９０６ｂと、決定された初期標準モデルを記憶する統計量記憶部９０６ｃと、統計量記憶部９０６ｃに記憶された

ることにより、参照モデル選択部９０５が選択した参照モデル９２３に対する確率又は尤度を最大化又は極大化するような統計量を算出する（最終的な標準モデルを生成する）統計量推定部９０６ｄとからなる。

音声認識部９１３は、標準モデル作成部９０６で作成された標準モデル９２２を用いて、マイク９１２から入力された利用者の音声を認識する。

次に、以上のように構成された携帯電話９０１の動作について説明する。

図３３は、携帯電話９０１の動作手順を示すフローチャートである。

いま、参照モデル記憶部９０３には、あらかじめ参照モデル９２１として子供用モデルが記憶されているとする。その参照モデル９２１は、音素ごとのＨＭＭにより構成される。参照モデル９２１の一例を図３４に示す。ここでは、子供用参照モデルのイメージ図が示されている。これらの参照モデルは、状態数３個、各状態は分布数が１６個の混合ガウス分布によりＨＭＭの出力分布が構成される。特徴量として、１２次元のメルケプストラム係数、１２次元のデルタメルケプストラム係数、デルタパワーの合計２５次元（Ｊ＝２５）の特徴量が用いられる。

まず、利用情報作成部９０４は、利用者の属するカテゴリである利用情報９２４を作成する（ステップＳ９００）。図３６は、利用情報９２４の作成例を示す図である。図３６（ａ）に携帯電話９０１の選択画面の一例を示す。ここでは、「４：成人」のボタンを押すことにより、この携帯電話９０１が成人女性と成人男性に利用されることが選択されている。別の一例を図３６（ｂ）に示す。ここでは、「メニュー」ボタンを押しながら音声を入力している。その利用者の音声は、特徴量に変換されることで、利用情報９２４である"利用者の音声データ"が作成される。

一方、仕様情報作成部９０７は、携帯電話９０１の仕様に基づき、仕

様情報９２５を作成する（ステップＳ９０１）。ここでは、携帯電話９０１のメモリ容量の大きさに基づいて「混合分布数１６」という仕様情報９２５を作成する。

次に、類似度情報作成部９０８は、利用情報９２４と仕様情報９２５と参照モデル記憶部９０３が記憶した参照モデル９２１に基づいて、類似度情報９２６を作成して（ステップＳ９０２）、類似度情報９２６を参照モデル準備部９０２に送信する。ここでは、参照モデル記憶部９０３に存在する参照モデル９２１は、混合分布数３の子供用モデル（図３４を参照）のみであり、利用情報９２４である「成人」（図３６（ａ）に対応）と仕様情報９２５である「混合分布数１６」に対応する参照モデルが参照モデル記憶部９０３に存在しないため、「類似した参照モデルが存在しない」という類似度情報９２６を作成して、類似度情報９２６を参照モデル準備部９０２に送信する。別の一例では、利用情報９２４は"利用者の音声データ"（図３６（ｂ）に対応）であり、利用者の音声データを参照モデル記憶部９０３が記憶している子供用モデルに入力して類似度情報９２６を作成する。ここでは、子供用モデルに対する尤度が所定のしきい値以下であるため、「類似した参照モデルが存在しない」という類似度情報９２６を作成して参照モデル準備部９０２に送信する。

続いて、参照モデル準備部９０２は、類似度情報９２６に基づいて、参照モデルを準備するか否かを決定する（ステップＳ９０３）。ここでは、「類似した参照モデルが存在しない」ため、図３７（ａ）の携帯電話９０１の画面表示例に示すように利用者に参照モデルの準備を促す。ここで、利用者が「メモ」ボタンを押して参照モデルの準備を要求した場合に、参照モデル準備部９０２は、参照モデルを準備すると決定して、利用情報９２４と仕様情報９２５を参照モデル受信部９０９に送信する。別の一例では、「類似した参照モデルが存在しない」ため、参照モデル準

備部９０２は、自動的に参照モデルを準備すると決定して、利用情報９２４と仕様情報９２５を参照モデル受信部９０９に送信する。この場合の携帯電話９０１の画面の一例を図３７（ｂ）に示す。

これに対して、参照モデル受信部９０９は、利用情報９２４と仕様情報９２５に対応した参照モデルをサーバ装置９１０から受信して参照モデル準備部９０２に送信する。ここでは、参照モデル受信部９０９は、利用情報９２４である「成人」（図３６（ａ）に対応）と仕様情報９２５である「混合分布数１６」に対応する参照モデルである、"混合分布数１６の成人女性用モデル"と"混合分布数１６の成人男性用モデル"の２個の参照モデルをサーバ装置９１０から受信する。

そして、参照モデル準備部９０２は、参照モデル受信部９０９が送信した参照モデルを参照モデル記憶部９０３に記憶することによって参照モデルを準備する（ステップＳ９０４）。図３５にその参照モデルの一例を示す。ここでは、成人男性用、成人女性用、子供用の参照モデルのイメージ図が示されている。

次に、参照モデル選択部９０５は、利用情報９２４である「成人」に対応した同じカテゴリに属する"混合分布数１６の成人女性用モデル"と"混合分布数１６の成人男性用モデル"の２個の参照モデルを参照モデル記憶部９０３が記憶している参照モデル９２１の中から選択する（ステップＳ９０５）。別の一例では、参照モデル選択部９０５は、利用情報９２４である"利用者の音声データ"と音響的に近い（尤度が大きい）"混合分布数１６の成人女性用モデル"と"混合分布数１６の成人男性用モデル"の２個の参照モデルを参照モデル記憶部９０３が記憶している参照モデル９２１の中から選択する。

続いて、標準モデル作成部９０６は、作成された仕様情報９２５に基づいて、参照モデル選択部９０５が選択した参照モデル９２３に対する

確率又は尤度を最大化又は極大化するように標準モデル９２２を作成する（ステップＳ９０６）。

最後に、音声認識部９１３は、標準モデル作成部９０６によって作成された標準モデル９２２に従って、マイク９１２から入力された利用者の音声を認識する（ステップＳ９０７）。

次に、図３３におけるステップＳ９０６（標準モデルの作成）の詳細な手順を説明する。手順の流れは、図４に示されたフローチャートと同様である。ただし、採用する標準モデルの構造や具体的な近似計算等が異なる。

まず、標準モデル構造決定部９０６ａは、標準モデルの構造を決定する（図４のステップＳ１０２）。ここでは、標準モデルの構造として、仕様情報９２５である「混合分布数１６」に基づいて、音素ごとのＨＭＭにより構成し、状態数を３とし、各状態における出力分布の混合分布数を１６個（Ｍｆ＝１６）と決定する。

次に、初期標準モデル作成部９０６ｂは、標準モデルを計算するための統計量の初期値を決定する（図４のステップＳ１０２ｂ）。ここでは、選択された参照モデル９２３である"混合分布数１６の成人女性用モデル"を統計量の初期値として統計量記憶部９０６ｃに記憶する。別の一例では、選択された参照モデル９２３である"混合分布数１６の成人男性女モデル"を統計量の初期値として統計量記憶部９０６ｃに記憶する。具体的には、初期標準モデル作成部９０６ｂは、上記式１３に示される出力分布を生成する。

そして、統計量推定部９０６ｄは、参照モデル選択部９０５が選択した２個の参照モデル９２３を用いて、統計量記憶部９０６ｃに記憶された標準モデルの統計量を推定する（図４のステップＳ１０２ｃ）。つまり、２個（Ｎｇ＝２）の参照モデル９２３における出力分布、即ち、上記式

１９に示される出力分布に対する標準モデルの確率（ここでは、上記式２５に示される尤度 logP）を極大化もしくは最大化するような標準モデルの統計量（上記式１６に示される混合重み係数、上記式１７に示される平均値、及び、上記式１８に示される分散値）を推定する。ただし、本実施の形態では、上記式１９に示された出力分布における式２１は、１６（各参照モデルの混合分布数）である。

具体的には、上記式２６、式２７及び式２８に従って、それぞれ、標準モデルの混合重み係数、平均値及び分散値を算出する。

このとき、統計量推定部９０６ｄの第３近似部９０６ｅは、標準モデルの各ガウス分布はお互いに影響を与えないと仮定して、式５３の近似式を用いる。また、繰り返し回数Ｒが１回目の場合には、式５４に示される標準モデルのガウス分布の近傍の式５５とは、式５４が示す出力分布とのマハラノビス距離、カルバック・ライブラー（ＫＬ）距離などの分布間距離が最も近いものと２番目に近いものの２個（近傍指示パラメータＧ＝２）の式５６に示される参照モデル９２３のガウス分布が存在する空間であると近似する。一方、繰り返し回数Ｒが２回目以上の場合には、式５４に示される標準モデルのガウス分布の近傍の式５５とは、式５４が示す出力分布とのマハラノビス距離、カルバック・ライブラー（ＫＬ）距離などの分布間距離が最も近いもの１個（近傍指示パラメータＧ＝１）の式５６に示される参照モデル９２３のガウス分布が存在する空間であると近似する。

以上の第３近似部９０６ｅによる近似式を考慮してまとめると、統計量推定部９０６ｄでの計算式は、次の通りになる。つまり、統計量推定部９０６ｄは、式５９、式６０及び式６１に従って、それぞれ、混合重み係数、平均値及び分散値を算出し、それらのパラメータによって特定される標準モデルを最終的な標準モデル９２２として生成する。ただし、

第３の実施の形態における第２の方法である、混合重み係数の値をゼロにして、平均値をゼロ、分散値を１にする方法を用いる。また、繰り返し回数に対応して近傍指示パラメータＧの値は異なる。なお、近傍指示パラメータＧの値に依存して、上記の方法を、第３の実施の形態における第１から第３の方法のいずれかに決定してもよい。

統計量推定部９０６ｄは、このように推定した標準モデルの統計量を統計量記憶部９０６ｃに記憶する。そして、このような統計量の推定と統計量記憶部９０６ｃへの記憶をＲ（≧１）回、繰り返す。その結果得られた統計量を最終的に生成する標準モデル９２２の統計量として出力する。

図３８に、第３近似部９０６ｅを用いて作成した標準モデル９２２を用いた認識実験の結果を示す。縦軸に成人（男性と女性）の認識率（％）、横軸に繰り返し回数Ｒを示す。繰り返し回数Ｒ＝０とは、学習を行う前での初期標準モデル作成部９０６ｂが作成した初期モデルにより認識した結果である。また、繰り返し回数Ｒ＝１のときは、近傍指示パラメータＧ＝２とし、繰り返し回数Ｒ＝２～５のときは、近傍指示パラメータＧ＝１とした。

グラフ「データ」は、数日間かけて音声データより学習した場合の結果を表しており、グラフ「女性」、グラフ「男性」は、それぞれ、初期モデルを成人女性、成人男性としたときの結果を表している。参照モデルによる本発明による学習時間は数十秒のオーダーであった。実験結果より、短時間に高い精度の標準モデルが作成できていることがわかる。

ここで、参考のために、図３９に、第３の実施の形態における第２近似部３０６ｅにより作成された標準モデルによる認識率を示す。本実施の形態における第３近似部９０６ｅと異なるのは、繰り返し回数Ｒによらず近傍指示パラメータＧ＝１であるということである。実験結果より、

初期モデルとして成人女性を選択すると良好な結果が得られることがわかる。また、初期モデルとして成人男性を選択すると、精度が少し劣化していることがわかる。図３８の結果とあわせると、第３近似部９０６ｅによる標準モデルは初期モデルに依存せずに高い精度の標準モデルが作成できていることがわかる。

以上説明したように、本発明の第８の実施の形態によれば、類似度情報に基づいて参照モデルを準備するため、利用情報及び仕様情報にふさわしい参照モデルを必要なタイミングで準備することができる。また、近傍指示パラメータＧを繰り返し回数Ｒによって変化させることで、初期モデルにかかわらず精度の高い標準モデルを提供することができる。

なお、統計量推定部９０６ｄによる処理の繰り返し回数は、上記式２５に示された尤度の大きさがある一定のしきい値以上になるまでの回数としてもよい。

また、標準モデル９２２は、音素ごとにＨＭＭを構成するに限らず、文脈依存のＨＭＭで構成してもよい。

また、標準モデル作成部９０６は、一部の音素の、一部の状態における事象の出力確率に対してモデル作成を行ってもよい。

また、標準モデル９２２を構成するＨＭＭは、音素ごとに異なる状態数により構成してもよいし、状態ごとに異なる分布数の混合ガウス分布により構成してもよい。

また、標準モデルを作成したのちに、さらに音声データにより学習してもよい。

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのＨＭＭの構造や、状態数などを決定してもよい。

全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がＰＤＡ（Ｐｅｒｓｏｎａｌ　Ｄｉｇｉｔａｌ　Ａｓｓｉｓｔａｎｔ）１００１に組み込まれた例が示されている。以下、本実施の形態では音声認識用の標準モデルを作成する場合を例にして説明する。

ＰＤＡ１００１は、携帯情報端末であり、事象の集合と事象又は事象間の遷移の出力確率で表現された隠れマルコフモデルによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、参照モデル記憶部１００３と、標準モデル作成部１００６と、アプリ・仕様情報対応データベース１０１４と、マイク１０１２と、音声認識部１０１３とを備える。標準モデル作成部１００６は、標準モデル構造決定部１００６ａと、初期標準モデル作成部１００６ｂと、統計量記憶部３０６ｃと、統計量推定部３０６ｄとを備える。

標準モデル作成部１００６は、送信されたアプリ起動情報１０２７（ここでは、起動したアプリケーションのＩＤ番号）に基づいて、アプリ・仕様情報対応データベース１０１４を用いて、仕様情報１０２５を取得する。図４１は、仕様情報対応データベース１０１４のデータ例を示す。仕様情報対応データベース１０１４には、アプリケーション（ＩＤ番号及び名前）に対応する仕様情報（ここでは、混合分布数）が登録されている。

標準モデル作成部１００６は、取得した仕様情報１０２５に基づいて、参照モデル記憶部１００３が記憶した１個の参照モデル１０２１に対する確率又は尤度を最大化又は極大化するように標準モデル１０２２を作成する処理部であり、第３の実施の形態における第２近似部３０６ｅの機能を有する。

音声認識部１０１３は、標準モデル作成部１００６で作成された標準

を認識する。

次に、以上のように構成されたＰＤＡ１００１の動作について説明する。

図４２は、ＰＤＡ１００１の動作手順を示すフローチャートである。

ここで、参照モデル記憶部１００３には、あらかじめ多くの混合分布数をもつ利用者用モデルが参照モデル１０２１として１個、記憶されているとする。参照モデル１０２１は、音素ごとのＨＭＭにより構成される。参照モデル１０２１の一例を図４３に示す。この参照モデルは、状態数３個、各状態は分布数が３００個の混合ガウス分布によりＨＭＭの出力分布が構成される。特徴量として、１２次元のメルケプストラム係数、１２次元のデルタメルケプストラム係数、デルタパワーの合計２５次元（Ｊ＝２５）の特徴量が用いられる。

まず、利用者は、例えば「株取引」というアプリケーションを起動する（ステップＳ１０００）。

これに対して、標準モデル作成部１００６は、アプリ起動情報として起動されたアプリケーションのＩＤ「３」を受信する（ステップＳ１００１）。そして、アプリ・仕様情報対応データベース１０１４を用いてＩＤ「３」に対応する仕様情報１０２５である「混合分布数１２６」に基づいて、標準モデル１０２２を作成する（ステップＳ１００２）。具体的には、標準モデル１０２２として、混合分布数１２６（Ｍｆ＝１２６）で、３状態の文脈依存型のＨＭＭにより構成する。

次に、標準モデル作成部１００６は、仕様情報１０２５を受信して（ステップＳ１００１）、仕様情報１０２５に基づいて標準モデルを作成する（ステップＳ１００２）。

最後に、音声認識部１０１３は、標準モデル作成部１００６によって作成された標準モデル１０２２に従って、マイク１０１２から入力され

た利用者の音声を認識する（ステップＳ１００３）。

次に、図４２におけるステップＳ１００２（標準モデルの作成）の詳細な手順を説明する。手順の流れは、図４に示されたフローチャートと同様である。ただし、採用する標準モデルの構造や具体的な近似計算等が異なる。

まず、標準モデル構造決定部１００６ａは、アプリ起動情報１０２７としてアプリケーションＩＤ「３」を受信した後に、アプリ・仕様情報対応データベース１０１４を用いてＩＤ「３」に対応した仕様情報１０２５（「混合分布数１２６」）を参照することにより、標準モデルの構造を混合分布数１２６（Ｍｆ＝１２６）で、３状態の文脈依存型のＨＭＭと決定する（図４のステップＳ１０２ａ）。

そして、初期標準モデル作成部１００６ｂは、標準モデル構造決定部１００６ａが決定した標準モデルの構造に基づいて、標準モデルを計算するための統計量の初期値を決定する（図４のステップＳ１０２ｂ）ここでは、k-means法とマハラノビス汎距離を用いた方法により、後述するクラスタリングを行ったものを統計量の初期値として統計量記憶部３０６ｃに記憶する。

そして、統計量推定部３０６ｄは、参照モデル記憶部１００３に格納された参照モデル１０２１を用いて、統計量記憶部３０６ｃに記憶された標準モデルの統計量を推定する（図４のステップＳ１０２ｃ）。なお、この統計量推定部３０６ｄによる推定処理は、第３の実施の形態と同様である。

次に、初期標準モデル作成部１００６ｂによる初期値の決定方法、つまり、k-means法とマハラノビス汎距離を用いた方法によるクラスタリングについて説明する。図４４にクラスタリングのフローチャートを示す。また、図４５～図４８にクラスタリングのイメージ図を示す。

まず、図４４のステップＳ１００４において、標準モデルの混合分布数である１２６個の代表点を準備する（図４５）。ここでは、参照モデルの３００個の出力分布の中から１２６個の出力分布を選択して、選択された分布の平均値を代表点とする。

次に、図４４のステップＳ１００５において、各代表点にマハラノビス汎距離が近い参照モデルの出力ベクトルを決定する(図４６)。そして、図４４のステップＳ１００６において、ステップＳ１００５で決定した近い分布を１つのガウス分布で表現して平均値を新しい代表点とする（図４７）。

続いて、図４４のステップＳ１００７において、クラスタリング操作を停止するかどうかを決定する。ここでは、各代表点と参照ベクトルの分布とのマハラノビス汎距離の変化率（１回前の代表点との距離との差分）がしきい値以下になった場合に停止とする。停止条件を満たさない場合、図４４のステップＳ１００５に戻り、近い分布を決定して同様の操作を繰り返す。

一方、停止条件を満たす場合には、図４４のステップＳ１００８に進み、統計量の初期値を決定して統計量記憶部３０６ｃに記憶する。このようにして、クラスタリングによる初期値の決定が行われる。

以上説明したように、本発明の第９の実施の形態によれば、アプリケーションに連動して自動的に仕様情報にふさわしい標準モデルを獲得することができる。

なお、標準モデル１０２２は、音素ごとにＨＭＭを構成してもよい。

また、標準モデル作成部１００６は、一部の音素の、一部の状態における事象の出力確率に対してモデル作成を行ってもよい。

また、標準モデル１０２２を構成するＨＭＭは、音素ごとに異なる状

布により構成してもよい。

また、標準モデルを作成したのちに、さらに音声データにより学習してもよい。

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのＨＭＭの構造や、状態数などを決定してもよい。

（第１０の実施の形態）

図４９は、本発明の第１０の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ８０１に組み込まれた例が示されている。本実施の形態では音声認識用の標準モデル（適応モデル）を作成する場合を例にして説明する。

サーバ８０１は、通信システムにおけるコンピュータ装置等であり、事象の集合と事象又は事象間の遷移の出力確率とによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、読み込み部７１１と、参照モデル準備部７０２と、参照モデル記憶部７０３と、利用情報受信部７０４と、参照モデル選択部７０５と、標準モデル作成部７０６と、仕様情報受信部７０７と、標準モデル記憶部７０８と、標準モデル送信部７０９と、参照モデル受信部８１０とを備える。

参照モデル準備部７０２は、読み込み部７１１が読み込んだ、ＣＤ－ＲＯＭなどのストレージデバイスに書き込まれた話者・雑音・声の調子別の音声認識用参照モデルを参照モデル記憶部７０３へ送信する。参照モデル記憶部７０３は、送信された参照モデル７２１を記憶する。また、参照モデル準備部７０２は、端末装置７１２からの送信に対して参照モデル受信部８１０が受信した音声認識用参照モデルを参照モデル記憶部７０３へ送信する。参照モデル記憶部７０３は、送信された参照モデル

仕様情報受信部７０７は、端末装置７１２から仕様情報７２５を受信する。利用情報受信部７０４は、端末装置７１２から利用情報７２４である雑音下で発声した利用者の音声を受信する。参照モデル選択部７０５は、利用情報受信部７０４が受信した利用情報７２４である利用者の音声に音響的に近い話者・雑音・声調子の参照モデル７２３を、参照モデル記憶部７０３が記憶している参照モデル７２１から選択する。

標準モデル作成部７０６は、仕様情報７２５に基づいて、参照モデル選択部７０５が選択した参照モデル７２３に対する確率又は尤度を最大化又は極大化するように標準モデル７２２を作成する処理部であり、第２の実施の形態における標準モデル作成部２０６と同一の機能を有する。標準モデル記憶部７０８は、仕様情報７２５に基づいた１もしくは複数の標準モデルを記憶する。標準モデル送信部７０９は、利用者の端末装置７１２から、仕様情報７２５と標準モデルの要求信号とを受信すると、その端末装置７１２へ、仕様に適した標準モデルを送信する。

次に、以上のように構成されたサーバ８０１の動作について説明する。

図５０は、サーバ８０１の動作手順を示すフローチャートである。なお、このサーバ８０１の動作手順を説明するための参照モデル及び標準モデルの一例は、第７に実施の形態における図３１と同様である。

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（図５０のステップＳ８００、Ｓ８０１）。つまり、参照モデル準備部７０２は、読み込み部７１１が読み込んだ、ＣＤ－ＲＯＭなどのストレージデバイスに書き込まれた話者・雑音・声の調子別の音声認識用参照モデルを参照モデル記憶部７０３へ送信し、参照モデル記憶部７０３は、送信された参照モデル７２１を記憶する（図５０のステップＳ８００）。ここでは、参照モデル７２１は、話者・雑音・声の調子ごとに、音

端末装置７１２が送信して参照モデル受信部８１０が受信した、利用者と端末装置７１２に適した音声認識用参照モデルを参照モデル記憶部７０３へ送信し、参照モデル記憶部７０３は、送信された参照モデル７２１を記憶する(図５０のステップＳ８０１)。ここでは、各参照モデルは、図３１の参照モデル７２１に示されるように、状態数３個、各状態は混合分布数が１２８個の混合ガウス分布によりＨＭＭの出力分布が構成される。特徴量として２５次元（Ｊ＝２５）のメルケプストラム係数が用いられる。

以下、これらの参照モデル７２１を用いた標準モデル７２２の作成及び端末装置７１２への送信(図５０のステップＳ８０２～Ｓ８０９)は、第７の実施の形態における手順(図３０のステップＳ７０１～Ｓ７０８）と同様である。

このようにして、端末装置７１２に記憶された自分用モデルをサーバにアップロードして標準モデル作成の材料にすることができるので、例えば、サーバ８０１において、アップロードされてきた参照モデルと既に保持している他の参照モデルとを統合して更に混合数の多い高精度の標準モデルを作成し、端末装置７１２にダウンロードして利用することが可能となる。したがって、端末装置７１２に簡易的な適応機能が付属され、簡易的に適応したモデルをアップロードして、さらに高精度な標準モデルを作成することもできる。

図５１は、本実施の形態における標準モデル作成装置を具体的に適用したシステム例を示す図である。ここには、インターネットや無線通信等を介して通信し合うサーバ７０１と端末装置７１２（携帯電話機７１２ａ、カーナビゲーション装置７１２ｂ）とが示されている。

たとえば、携帯電話機７１２ａは、利用者の音声を利用情報とし、携

とし、予め記憶しているサンプルモデルを参照モデルとし、それら利用情報、仕様情報及び参照モデルをサーバ７０１に送信することで、標準モデルの作成を要求する。その要求に対してサーバ７０１で標準モデルが作成されると、携帯電話機７１２ａは、その標準モデルをダウンロードし、その標準モデルを用いて利用者の音声を認識する。例えば、利用者の音声が、内部に保持するアドレス帳の名前と一致した場合には、その名前に対応する電話番号に自動発呼する。

また、カーナビゲーション装置７１２ｂは、利用者の音声を利用情報とし、カーナビゲーション装置での利用である旨（ＣＰＵの処理能力が通常であること）を仕様情報とし、予め記憶しているサンプルモデルを参照モデルとし、それら利用情報、仕様情報及び参照モデルをサーバ７０１に送信することで、標準モデルの作成を要求する。その要求に対してサーバ７０１で標準モデルが作成されると、カーナビゲーション装置７１２ｂは、その標準モデルをダウンロードし、その標準モデルを用いて利用者の音声を認識する。例えば、利用者の音声が、内部に保持する地名と一致した場合には、その地名を目標点とする現地点からの道順を示す地図を画面に自動表示する。

このようにして、携帯電話機７１２ａ及びカーナビゲーション装置７１２ｂは、自装置に適した標準モデルの作成をサーバ７０１に依頼することで、標準モデルの作成に必要な回路や処理プログラムを自装置内に実装する必要がなくなるとともに、様々な認識対象の標準モデルを必要なタイミングで獲得することができる。

以上説明したように、本発明の第１０の実施の形態によれば、参照モデル受信部８１０が受信した参照モデルを利用して標準モデルを作成できるため、精度の高い標準モデルが提供される。つまり、端末装置７１

側で保持する参照モデルのバリエーションが増加し、他の人が利用したときにさらに高精度の標準モデルを提供することができる。

また、仕様情報に基づいて標準モデルが作成されるため、標準モデルを利用する機器にふさわしい標準モデルが準備される。

なお、参照モデル受信部８１０は、端末装置７１２とは異なる他の端末装置から参照モデルを受信してもよい。

また、図５１に示された応用例は、本実施の形態に限られるものではなく、他の実施の形態にも適用することができる。つまり、第１～第９の実施の形態で作成された標準モデルを各種記録媒体や通信を介して様々な電子機器に配信することで、それらの電子機器において、制度の高いな音声認識、画像認識、意図理解等を行うことが可能となる。さらに、上記実施の形態における標準モデル作成装置を各種電子機器に内蔵させることで、音声認識、画像認識、意図理解等の認識・認証機能を備えるスタンドアローンの電子機器を実現することもできる。

以上、本発明に係る標準モデル作成装置について、実施の形態に基づいて説明したが、本発明は、これらの実施の形態に限定されるものではない。

たとえば、第１～第１０の実施の形態における標準モデルの統計量の近似計算については、各実施の形態における近似計算だけに限られず、第１～第４の実施の形態における合計４種類の近似計算の少なくとも１つを用いてもよい。つまり、４種類の近似計算のいずれであってもよいし、２以上の種類の近似計算の組み合わせであってもよい。

また、第２の実施の形態では、統計量推定部２０６ｄの一般近似部２０６ｅは、標準モデルの混合重み係数、平均値及び分散値を、それぞれ、式４５、式４６及び式４７に示される近似式に従って算出したが、これ

式を用いて算出してもよい。

（式６３）

$$\omega_{f(m)} \approx \frac{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \{\sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, m) \upsilon_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma^2_{g(i,l)})\} dx}{\sum_{k=1}^{M_f} \sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \{\sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, k) \upsilon_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma^2_{g(i,l)})\} dx}$$

$$(m = 1,2,..., M_f)$$

（式６４）

$$\mu_{f(m,j)} \approx \frac{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} x_{(j)} \{\sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, m) \upsilon_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma^2_{g(i,l)})\} dx}{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \{\sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, m) \upsilon_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma^2_{g(i,l)})\} dx}$$

$$(m = 1,2,..., M_f, j = 1,2,..., J)$$

（式６５）

$$\sigma^2_{f(m,j)} \approx \frac{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} (x_{(j)} - \mu_{f(m,j)})^2 \{\sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, m) \upsilon_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma^2_{g(i,l)})\} dx}{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \{\sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, m) \upsilon_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma^2_{g(i,l)})\} dx}$$

$$(m = 1,2,..., M_f, j = 1,2,..., J)$$

このような近似式を用いて作成した標準モデルによれば、高い認識性能が得られることが発明者らによって確認されている。たとえば、参照モデルと標準モデルそれぞれの混合数を１６とした場合の認識結果は、

適応前では８２．２％であったものが、上記非特許文献２に示された十分統計量による方法では、８５．０％、上記近似式による方法では８５．５％に改善された。つまり、十分統計量による方法と比べ、高い認識性能が獲得できていることがわかる。また、参照モデルの混合数を６４、標準モデルの混合数を１６とした場合についての認識結果は、上記近似式による方法では、８５．７％と高い認識率が獲得できている。

また、初期標準モデル作成部による初期標準モデルの作成においては、図５２に示されるようなクラスＩＤ・初期標準モデル・参照モデル対応表を予め準備しておき、この表に従って、初期標準モデルを決定してもよい。以下、このようなクラスＩＤ・初期標準モデル・参照モデル対応表を用いた初期標準モデルの決定方法について説明する。なお、クラスＩＤとは、標準モデルを用いた認識対象の種別を識別するＩＤであり、標準モデルの種類に対応する。

図５２に示されたクラスＩＤ・初期標準モデル・参照モデル対応表は、一定の共通する性質を有する複数の参照モデルに対して、それらを識別する１つのクラスＩＤを対応づけるとともに、それら参照モデルと共通する性質を持つ予め作成された初期標準モデルを対応づけた表である。この表では、参照モデル８ＡＡ～８ＡＺに対して、クラスＩＤ及び初期標準モデル８Ａが対応づけられ、参照モデル６４ＺＡ～ＺＺに対して、クラスＩＤ及び初期標準モデル６４Ｚが対応づけられている。標準モデル作成部は、使用する参照モデルの性質と共通する初期標準モデルを使用することによって、精度の高い標準モデルを生成することができる。

ここで、クラスＩＤ、初期標準モデル及び参照モデルの添え字記号８Ａ、８ＡＡにおける最初の記号「８」等は、混合分布数を意味し、２番目の記号「Ａ」等は大分類、例えば、騒音下における音声認識の場合であれば、騒音環境の種類（家庭内騒音下をＡ、電車内騒音下をＢなど）

を意味し、３番目の記号「Ａ」等は小分類、例えば、音声認識の対象となる人の属性（低学年の小学生をＡ、高学年の小学生をＢなど）を意味する。したがって、図５２のクラスＩＤ・初期標準モデル・参照モデル対応表における参照モデル８ＡＡ～ＡＺは、図５３に示されるような混合分布数８のモデルであり、参照モデル６４ＺＡ～ＺＺは、図５４に示されるような混合分布数６４のモデルであり、初期標準モデル８Ａ～６４Ｚは、図５５に示されるような混合分布数８～１６のモデルである。

次に、このようなクラスＩＤ・初期標準モデル・参照モデル対応表の作成方法を説明する。図５６は、その手順を示すフローチャートであり、図５７～図６０は、各ステップでの具体例を示す図である。ここでは、騒音環境下での音声認識を例とし、表だけでなく、クラスＩＤ、初期標準モデル及び参照モデルも含めて新規に作成する場合の手順を説明する。

まず、音声データを音響的に近いグループに分類する（図５６のステップＳ１１００）。たとえば、図５７に示されるように、音声データを利用情報である雑音環境で分類する。環境Ａ（家庭内騒音下での音声データ）には、家庭内騒音下で収録した小学生低学年の音声、小学生高学年の音声、成人女性の音声などが含まれ、環境Ｂ（電車内での音声データ）には、電車内で収録した小学生低学年の音声、小学生高学年の音声、成人女性の音声などが含まれるように分類する。なお、利用情報である話者の性別、年齢層、笑い声・怒った声などの声の性質、読み上げ調・会話調などの声の調子、英語・中国語などの言語などで分類してもよい。

次に、仕様情報等に基づいて、準備する参照モデルの１以上のモデル構造を決定する（図５６のステップＳ１１０１）。たとえば、８混合、１６混合、３２混合及び６４混合を対象とすることを決定する。なお、モデル構造の決定においては、混合分布数を決定するに限らず、ＨＭＭの

もよい。

続いて、初期標準モデルを作成する（図５６のステップＳ１１０２）。つまり、上記音声データの分類（ステップＳ１１００）において決定した分類（環境Ａ、環境Ｂ、…）ごとに、ステップＳ１１０１において決定したモデル構造ごとの初期標準モデルを作成する。例えば、図５８に示されるように、初期標準モデル８Ａであれば、８混合の初期標準モデルを、家庭内騒音下（環境Ａ）における音声データ（低学年の小学生、高学年の小学生、成人男、成人女等の音声データ）を用いて、バウム・ウェルチアルゴリズムなどにより学習して作成する。

次に、参照モデルを作成する（図５６のステップＳ１１０３）。つまり、上記ステップＳ１１０２において作成した初期標準モデルを用いて参照モデルを作成する。具体的には、参照モデルを学習する音声データの雑音環境と同じ雑音環境で学習した、同じ混合分布数をもつ初期標準モデルを用いて参照モデルを学習する。例えば、図５９に示されるように、参照モデル８ＡＡは、混合分布数８の家庭内騒音下での小学生低学年の音声データで学習するモデルであり、学習を行う際の初期値として、同じ環境である家庭内騒音下での音声データ（小学生低学年、小学生高学年、成人女性、成人男性の音声を含む）で学習した初期標準モデルを用いる。学習方法として、バウム・ウェルチアルゴリズムを用いる。

最後に、クラスＩＤを付与する（図５６のステップＳ１１０４）。たとえば、騒音環境下ごとに１つのクラスＩＤを付与することによって、図６０に示されるクラスＩＤ・初期標準モデル・参照モデル対応表、つまり、"クラスＩＤ付き初期標準モデル"及び"クラスＩＤ付き参照モデル"が作成される。

なお、このようなクラスＩＤ・初期標準モデル・参照モデル対応表は、

要はない。端末（標準モデル作成装置）は、図６１に示されるように、他の装置（サーバ）と通信することによって表を完成させてもよい。つまり、標準モデル作成装置（端末）は、通信網などを介して、"クラスＩＤ付き初期標準モデル", "クラスＩＤ付き参照モデル"を取得することが可能である。もっとも、端末は必ずしも"クラスＩＤ付き初期標準モデル"、"クラスＩＤ付参照モデル"を取得する必要はなく事前に記憶させて出荷してもよい。

図６１に示されるように、端末は、以下のような方法によって、"クラスＩＤ付き初期標準モデル"、"クラスＩＤ付き参照モデル"を取得することができる。第１の方法として、端末は、"クラスＩＤ付き初期標準モデル"（例えば規格化コンソーシアムなどで事前に定義されたクラスＩＤのつけ方に遵守したもの）を記憶しているケースである。このとき、端末は、１以上のサーバから"クラスＩＤ付き参照モデル"（例えば規格化コンソーシアムなどで事前に定義されたクラスＩＤのつけ方に遵守したもの）をダウンロードする。なお、端末に、"クラスＩＤ付き参照モデル"を出荷時に記憶させておいてもよい。

また、第２の方法として、端末は、"クラスＩＤ付き初期標準モデル"を記憶していないケースである。このとき、端末は、サーバ（図６１のサーバ１）から"クラスＩＤ付き初期標準モデル"をダウンロードする。次に、端末は、１以上のサーバ（図６１のサーバ２）から"クラスＩＤ付き参照モデル"をダウンロードする。必要に応じて逐次的にクラスＩＤの定義の追加、変更が可能である。また、端末のメモリの節約にもなる。

さらに、第３の方法として、端末は、クラスＩＤと初期標準モデル・参照モデルの対応関係を明記した"クラスＩＤ・初期標準モデル・参照モデル対応表"を記憶しているケースである。このとき、端末は、"対応表"記憶していないサーバ（図６１のサーバ３）に"対応表"をアップロード

する。サーバは、送信された"対応表"に基づき"クラスＩＤ付き参照モデル"を準備する。端末は、準備された"クラスＩＤ付き参照モデル"をダウンロードする。

次に、このようなクラスＩＤ・初期標準モデル・参照モデル対応表を用いた初期標準モデル作成部による初期標準モデルの決定方法について説明する。図６２は、その手順を示すフローチャートである。図６３及び図６４は、各ステップでの具体例を示す図である。

まず、標準モデルの作成に用いる参照モデルからクラスＩＤを抽出する（図６２のステップＳ１１０５）。たとえば、図６３に示されるテーブルに従って、選択された参照モデルから、対応するクラスＩＤを抽出する。ここでは、抽出したクラスＩＤとして、８Ａが１個、１６Ａが３個、１６Ｂが１個、６４Ｂが１個とする。

次に、抽出したクラスＩＤを用いて標準モデル作成に用いる初期標準モデルを決定する（図６２のステップＳ１１０６）。具体的には、以下の手順に従って初期標準モデルを決定する。

（１）作成する標準モデルの混合分布数（１６混合）と同じクラスＩＤ（１６＊）をもつ参照モデルから抽出したクラスＩＤ（１６Ａ、１６Ｂ）に着目し、その中から一番多く抽出されたクラスＩＤに対応する初期標準モデルを最終的な初期標準モデルと決定する。たとえば、標準モデルの構造が１６混合の場合には、１６混合に関するクラスＩＤとして、１６Ａが３個、１６Ｂが１個抽出されているので、クラスＩＤが１６Ａの初期標準モデルを採用する。

（２）作成する標準モデルの混合分布数（８混合）と同じクラスＩＤ（８＊）をもつ参照モデルから抽出したクラスＩＤ（８Ａ）に着目し、同じクラスＩＤをもつ初期標準モデルを最終的な初期標準モデルと決定する。

スＩＤとして、８Ａが１個抽出されているので、クラスＩＤが８Ａの初期標準モデルを採用する。

（３）作成する標準モデルの混合分布数（３２混合）と同じクラスＩＤ（３２＊）をもつ参照モデルから抽出したクラスＩＤに着目し、存在しない場合、仕様情報に着目してその中から一番多く抽出されたクラスＩＤ（＊Ａ）をもつ初期標準モデル（８Ａ、１６Ａ）を用いてクラスタリングにより３２混合にして最終的な初期標準モデルとする（図４４を参照）。たとえば、標準モデルの構造が３２混合の場合には、３２混合に関するクラスＩＤが抽出されていないので、一番多く抽出されたクラスＩＤ（１６Ａ）を用いてクラスタリングにより３２混合にして初期標準モデルとする。

なお、はじめに作成する標準モデルの仕様情報（混合分布数など）に着目せず、利用情報（雑音の種類など）に着目して初期値を決定してもよい。

図６４に、第３近似部を用いて作成した混合分布数が６４の標準モデルを用いた認識実験の結果を示す。縦軸に成人（男性と女性）の認識率（％）、横軸に繰り返し回数Ｒを示す。繰り返し回数Ｒ＝０とは、学習を行う前での初期標準モデル作成部が作成した初期モデルにより認識した結果である。また、繰り返し回数Ｒ＝１～５において、近傍指示パラメータＧ＝１とした。

グラフ「データ」は、数日間かけて音声データより学習した場合の結果を表しており、グラフ「女性」、グラフ「男性」は、それぞれ、初期モデルを成人女性、成人男性としたときの結果を表している。参照モデルによる本発明による学習時間は数分のオーダーであった。この実験結果より、成人女性の参照モデルを初期標準モデルと決定した場合には、音声データで学習した結果よりも高い精度の標準モデルが作成できている

ことが分かる。

このことは、音声データを分割し、分割した音声データをそれぞれの参照モデルとして厳密に学習したのちに統合したほうが、音声データによる学習の課題である局所解に陥るという問題を解決できる可能性を示している（音声データによる学習との認識精度での比較）。

また、音声データの収録が困難な子供の音声データに対しては、データ数に適切である混合分布数の少ない参照モデルで厳密に学習して、多くの音声データの収録が可能な成人の音声データに対しては、混合分布数の多い参照モデルで厳密に学習して、そのあとで本発明により統合して標準モデルを作成すれば、極めて精度の高い標準モデルが作成できることが期待できる。

なお、標準モデルの混合分布数が１６の場合における認識実験（図３９）では、本発明による方法は、音声データで学習した標準モデルの認識率を超えていない。このことは、音声データを１６混合の参照モデルの形にしたときに音声データの情報が欠如したためだと考えられる。参照モデルを６４混合で作成して音声データの特徴を十分保持しておけばより高い精度の標準モデルが作成できる。このことより、第９の実施の形態では、参照モデルの混合分布数を３００と大きめに設定している。

また、図３９及び図６４に示される認識実験より、初期標準モデルが認識精度に与える影響が示されており、初期標準モデルの決定方法の重要性を物語っている（図６４において、成人女性の参照モデルを初期標準モデルとして利用した場合、成人男性の参照モデルを利用する場合より高い精度の標準モデルが作成できることが示されている）。

以上のように、クラスＩＤ・初期標準モデル・参照モデル対応表に従って、参照モデルと共通する性質の初期標準モデルを用いることで、精度の高い標準モデルを作成することができる。

なお、このようなクラスＩＤ・初期標準モデル・参照モデル対応表を用いた初期標準モデルの決定は、上記実施の形態１～１０のいずれにおいても採用することができる。

また、上記実施の形態では、標準モデルの統計量を推定する際に、参照モデルに対する標準モデルの尤度として式２５が用いられたが、本発明はこのような尤度関数に限られず、例えば、以下の式６６に示される尤度関数を用いてもよい。

（式６６）

$$\log L=\sum_{i=1}^{N}\int_{-\infty}^{\infty}\log\left[\sum_{m=1}^{M}\omega_{(m)}f(x;\mu_{(m)},\sigma^2_{(m)})\right]^{\alpha(i)}\left\{\sum_{l=1}^{L_i}\upsilon_{(l)}g_i(x;\mu_{(l)},\sigma^2_{(l)})\right\}dx$$

ここで、α(i)は、統合する各参照モデルｉに対応した重要度を示す重み付けである。たとえば、音声認識における話者適用であれば、重要度は、利用者の音声と統合モデルを作成した音声の近さにより決定される。つまり、参照モデルが利用者の音声に近い（重要度が大きい）場合に、α(i)は大きな値に設定される（大きく重み付けされる）。統合モデルと利用者の音声との近さは、利用者の音声を統合モデルに入力したときの尤度の大きさにより決定すればよい。これによって、複数の参照モデルを統合して標準モデルを作成する際に、利用者の音声に近い参照モデルほど大きな重み付けで標準モデルの統計量に影響を与えることとなり、より利用者の特性を反映した精度の高い標準モデルが作成される。

また、各実施の形態における標準モデル構造決定部は、利用情報や仕様情報などの各種要因に基づいて標準モデルの構造を決定したが、本発明は、これらの要因だけに限られず、例えば、音声認識の場合であれば、認識の対象となる人の年齢、性別、声質の話者性、感情又は健康状態に基づく声の調子、発話速度、発話の丁寧さ、方言、背景雑音の種類、背景雑音の大きさ、音声と背景雑音とのSN比、マイク特性及び認識語彙

の複雑さなどの各種属性に依存して標準モデルの構造を決定してもよい。

具体的には、図６５（ａ）～（ｊ）に示されるように、音声認識の対象となる人の年齢が高いほど標準モデルを構成するガウス分布の数（混合数）を大きくしたり（図６５（ａ））、音声認識の対象となる人が男性の場合には女性の場合よりも大きな混合数にしたり（図６５（ｂ））、音声認識の対象となる人の音質が「通常」よりも「ハスキー」、さらに「しわがれ声」となるほど混合数を大きくしたり（図６５（ｃ））、音声認識の対象となる声の感情による調子が「通常」よりも「怒り声」、さらに「泣き／笑いながらの声」となるほど混合数を大きくしたり（図６５（ｄ））、音声認識の対象となる人の発話速度が速く／遅くなるほど混合数を大きくしたり（図６５（ｅ））、音声認識の対象となる人の発話の丁寧さが「朗読調」よりも「講演調」、さらに「会話調」となるほど混合数を大きくしたり（図６５（ｆ））、音声認識の対象となる人の方言が「標準語」よりも「大阪弁」、さらに「鹿児島弁」となるほど混合数を大きくしたり（図６５（ｇ））、音声認識における背景雑音が大きくなるほど混合数を小さくしたり（図６５（ｈ））、音声認識に使用するマイクの性能が高くなるほど混合数を大きくしたり（図６５（ｉ））、音声認識の対象となる語彙が増加するほど混合数を大きくしたり（図６５（ｊ））すればよい。これらの例の多くは、認識対象の音声のばらつきが大きいほど、混合数を大きくして精度を確保するという観点から混合数が決定される。

産業上の利用の可能性

本発明に係る標準モデル作成装置は、確率モデル等を用いた音声、文字、画像等の対象物を認識する装置等として利用することができ、例えば、音声によって各種処理を実行するテレビ受信装置・カーナビゲーション装置、音声を他の言語に翻訳する翻訳装置、音声で操作するゲーム

装置、音声による検索キーワードで情報を検索する検索装置、人物検出・指紋認証・顔認証・虹彩認証等を行う認証装置、株価予測、天気予測などの予測を行う情報処理装置等として利用することができる。

請　求　の　範　囲

1．音声の特徴を示す周波数のパラメータを出力確率で表現する確率モデルを用いて、特定の属性を有する音声の特徴を示す音声認識用の標準モデルを作成する装置であって、

一定の属性を有する音声の特徴を示す確率モデルである１以上の参照モデルを記憶する参照モデル記憶手段と、

前記参照モデル記憶手段に格納された１以上の参照モデルの統計量を用いて前記標準モデルの統計量を計算することによって標準モデルを作成する標準モデル作成手段とを備え、

前記標準モデル作成手段は、

作成する標準モデルの構造を決定する標準モデル構造決定部と、

構造が決定された標準モデルを特定する統計量の初期値を決定する初期標準モデル作成部と、

初期値が決定された標準モデルの前記参照モデルに対する確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を推定して計算する統計量推定部とを有する

ことを特徴とする標準モデル作成装置。

2．前記標準モデル作成装置はさらに、

音声認識の対象となる属性に関する情報である利用情報に基づいて、前記参照モデル記憶手段に記憶されている参照モデルの中から１以上の参照モデルを選択する参照モデル選択手段を備え、

前記標準モデル作成手段は、前記参照モデル選択手段が選択した参照モデルの統計量を用いて標準モデルを作成する

ことを特徴とする請求の範囲１記載の標準モデル作成装置。

3．前記標準モデル作成装置はさらに、

前記利用情報を作成する利用情報作成手段を備え、

前記参照モデル選択手段は、作成された利用情報に基づいて、前記参照モデル記憶手段に記憶されている参照モデルの中から１以上の参照モデルを選択する

ことを特徴とする請求の範囲２記載の標準モデル作成装置。

4．前記標準モデル作成装置には通信路を介して端末装置が接続され、

前記標準モデル作成装置はさらに、

前記端末装置から前記利用情報を受信する利用情報受信手段を備え、

前記参照モデル選択手段は、受信された利用情報に基づいて、前記参照モデル記憶手段に記憶されている参照モデルの中から１以上の参照モデルを選択する

ことを特徴とする請求の範囲２記載の標準モデル作成装置。

5．前記標準モデル構造決定部は、作成する標準モデルの仕様に関する情報である仕様情報、及び、音声認識の対象となる属性に関する情報である利用情報の少なくとも一方に基づいて、前記標準モデルの構造を決定する

ことを特徴とする請求の範囲１記載の音声認識用の標準モデル作成装置。

6．前記仕様情報とは、標準モデルを使用するアプリケーションプログラムの種類、及び、標準モデルを使用する機器の仕様の少なくとも一方の仕様を示す

ことを特徴とする請求の範囲５記載の音声認識用の標準モデル作成装

置。

７．前記属性とは、年齢、性別、声質の話者性、感情又は健康状態に基づく声の調子、発話速度、発話の丁寧さ、方言、背景雑音の種類、背景雑音の大きさ、音声と背景雑音とのSN比、マイク特性及び認識語彙の複雑さの少なくとも１つに関する情報を含む

ことを特徴とする請求の範囲５記載の音声認識用の標準モデル作成装置。

８．前記標準モデル作成装置はさらに、

標準モデルを使用するアプリケーションプログラムと標準モデルの仕様との対応を示すアプリケーション仕様対応データベースを前記仕様情報として保持する仕様情報保持手段を備え、

前記標準モデル構造決定部は、前記仕様情報保持手段に保持されたアプリケーション仕様対応データベースから、起動されるアプリケーションプログラムに対応する仕様を読み出し、読み出した仕様に基づいて、前記標準モデルの構造を決定する

ことを特徴とする請求の範囲５記載の標準モデル作成装置。

９．前記標準モデル作成装置はさらに、

前記仕様情報を作成する仕様情報作成手段を備え、

前記標準モデル構造決定部は、作成された仕様情報に基づいて、前記標準モデルの構造を決定する

ことを特徴とする請求の範囲５記載の標準モデル作成装置。

１０．前記標準モデル作成装置には通信路を介して端末装置が接続され、

前記標準モデル作成装置はさらに、

前記端末装置から前記仕様情報を受信する仕様情報受信手段を備え、

前記標準モデル構造決定部は、受信された仕様情報に基づいて、前記標準モデルの構造を決定する

ことを特徴とする請求の範囲５記載の標準モデル作成装置。

１１．前記参照モデル及び前記標準モデルは、１以上のガウス分布を用いて表現され、

前記標準モデル構造決定部は、前記標準モデルの構造として、少なくともガウス分布の混合数を決定する

ことを特徴とする請求の範囲５記載の標準モデル作成装置。

１２．前記標準モデル作成装置には、通信路を介して端末装置が接続され、

前記標準モデル作成装置はさらに、

前記標準モデル作成手段が作成した標準モデルを前記端末装置に送信する標準モデル送信手段を備える

ことを特徴とする請求の範囲１記載の標準モデル作成装置。

１３．前記参照モデル及び前記標準モデルは、１以上のガウス分布を用いて表現され、

前記参照モデル記憶手段は、少なくともガウス分布の混合数が異なる１対の参照モデルを記憶し、

前記統計量推定部は、前記１対の参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を計算する

ことを特徴とする請求の範囲１記載の標準モデル作成装置。

１４．前記標準モデル作成手段はさらに、

外部から参照モデルを取得して前記参照モデル記憶手段に格納すること、及び、新たな参照モデルを作成して前記参照モデル記憶手段に格納することの少なくとも一方を行う参照モデル準備手段を備える

ことを特徴とする請求の範囲１記載の標準モデル作成装置。

１５．前記参照モデル準備手段は、さらに、前記参照モデル記憶手段が記憶する参照モデルの更新及び追加の少なくとも一方を行う

ことを特徴とする請求の範囲１４記載の標準モデル作成装置。

１６．前記参照モデル準備手段は、認識の対象に関する情報である利用情報、及び作成する標準モデルの仕様に関する情報である仕様情報の少なくとも一方に基づいて、前記参照モデル記憶手段が記憶する参照モデルの更新及び追加の少なくとも一方を行う

ことを特徴とする請求の範囲１５記載の標準モデル作成装置。

１７．前記標準モデル作成装置は、さらに、作成する標準モデルの仕様に関する情報である仕様情報、及び、音声認識の対象となる属性に関する情報である利用情報の少なくとも一方と、前記参照モデル記憶手段に記憶された参照モデルとに基づいて、前記利用情報及び前記仕様情報の少なくとも一方と前記参照モデルとの類似度を示す類似度情報を作成する類似度情報作成手段を備え、

前記参照モデル準備手段は、前記類似度情報作成手段が作成した類似度情報に基づいて、前記参照モデル記憶手段が記憶する参照モデルの更

新及び追加の少なくとも一方を行うか否かを決定する

ことを特徴とする請求の範囲１５記載の標準モデル作成装置。

１８．前記初期標準モデル作成部は、前記統計量推定部が標準モデルの統計量を計算するために用いる、１以上の前記参照モデルを用いて前記標準モデルを特定する統計量の初期値を決定する

ことを特徴とする請求の範囲１記載の標準モデル作成装置。

１９．前記初期標準モデル作成部は、標準モデルの種類を識別するクラスＩＤに基づいて、前記初期値を決定する

ことを特徴とする請求の範囲１記載の標準モデル作成装置。

２０．前記初期標準モデル作成部は、前記参照モデルから前記クラスＩＤを特定し、特定したクラスＩＤに対応づけられた初期値を前記初期値と決定する

ことを特徴とする請求の範囲１９記載の標準モデル作成装置。

２１．前記初期標準モデル作成部は、前記クラスＩＤと前記初期値と前記参照モデルとの対応を示す対応表を保持し、前記対応表に従って、前記初期値を決定する

ことを特徴とする請求の範囲２０記載の標準モデル作成装置。

２２．前記初期標準モデル作成部は、前記クラスＩＤが対応づけられた初期値であるクラスＩＤ付き初期標準モデル、又は、前記クラスＩＤが対応づけられた参照モデルであるクラスＩＤ付き参照モデルを作成又は外部から取得することによって、前記対応表を生成する

ことを特徴とする請求の範囲２１記載の標準モデル作成装置。

２３．前記参照モデル記憶手段は、複数の参照モデルを記憶し、

前記統計量推定部は、前記参照モデル記憶手段に記憶された複数の参照モデルに対して重み付けられた前記確率又は尤度を最大化又は極大化するように前記統計量を計算する

ことを特徴とする請求の範囲１記載の標準モデル作成装置。

２４．音声の特徴を示す周波数のパラメータを出力確率で表現する確率モデルを用いて、特定の属性を有する音声の特徴を示す音声認識用の標準モデルを作成する方法であって、

一定の属性を有する音声の特徴を示す確率モデルである１以上の参照モデルを記憶する参照モデル記憶手段から１以上の参照モデルを読み出す参照モデル読み出しステップと、

読み出された参照モデルの統計量を用いて前記標準モデルの統計量を計算することによって標準モデルを作成する標準モデル作成ステップとを含み、

前記標準モデル作成ステップは、

作成する標準モデルの構造を決定する標準モデル構造決定サブステップと、

構造が決定された標準モデルを特定する統計量の初期値を決定する初期標準モデル作成サブステップと、

初期値が決定された標準モデルの前記参照モデルに対する確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を推定して計算する統計量推定サブステップとを有する

ことを特徴とする標準モデル作成方法。

２５．音声の特徴を示す周波数のパラメータを出力確率で表現する確率モデルを用いて、特定の属性を有する音声の特徴を示す音声認識用の標準モデルを作成する装置のためのプログラムであって、

一定の属性を有する音声の特徴を示す確率モデルである１以上の参照モデルを記憶する参照モデル記憶手段から１以上の参照モデルを読み出す参照モデル読み出しステップと、

読み出された参照モデルの統計量を用いて前記標準モデルの統計量を計算することによって標準モデルを作成する標準モデル作成ステップとを含み、

前記標準モデル作成ステップは、

作成する標準モデルの構造を決定する標準モデル構造決定サブステップと、

構造が決定された標準モデルを特定する統計量の初期値を決定する初期標準モデル作成サブステップと、

初期値が決定された標準モデルの前記参照モデルに対する確率又は尤度を最大化又は極大化するように前記標準モデルの統計量を推定して計算する統計量推定サブステップとを有する

ことを特徴とするプログラム。

図1

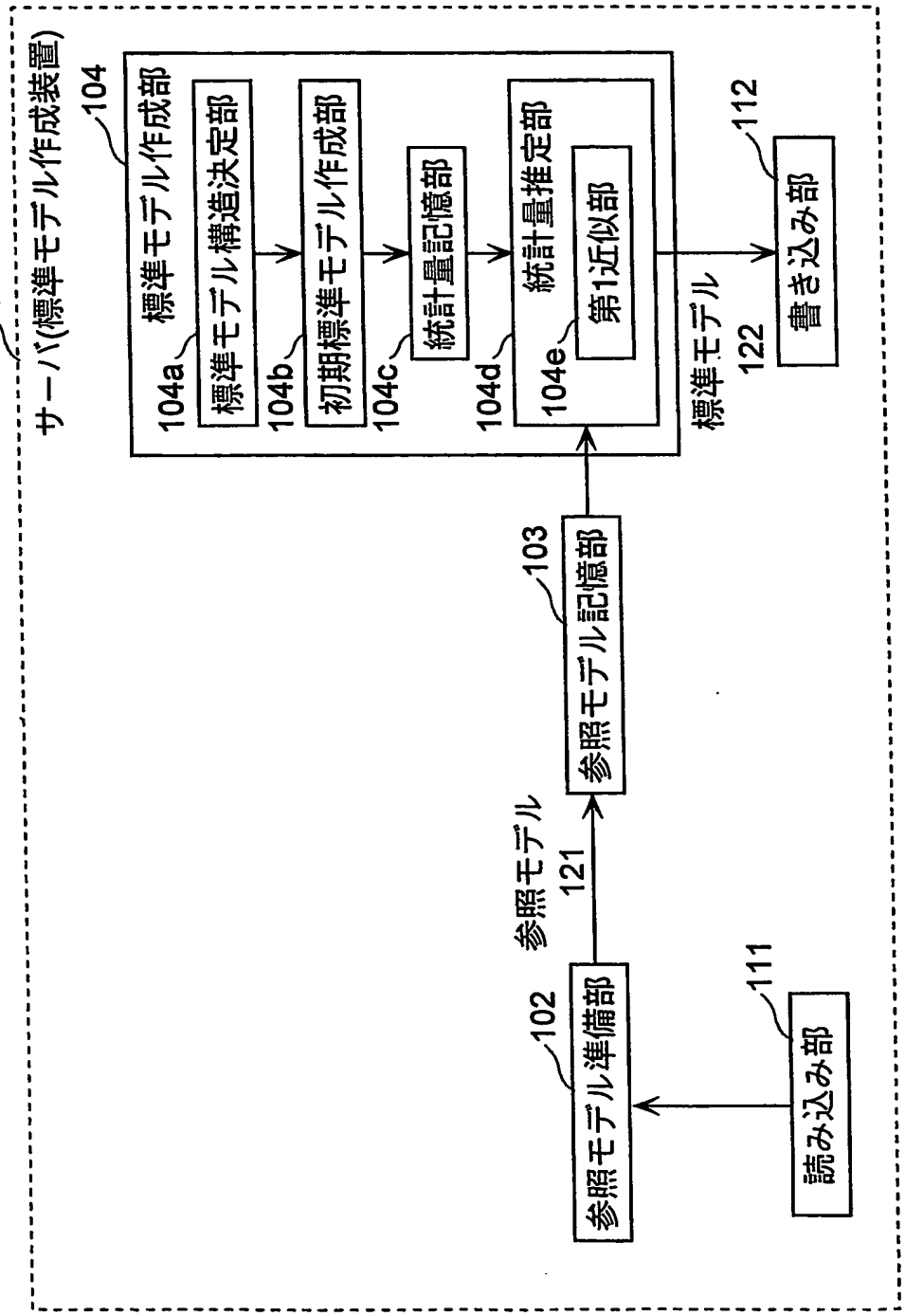

101
サーバ(標準モデル作成装置)
104
標準モデル作成部
104a
標準モデル構造決定部
104b
初期標準モデル作成部
104c
統計量記憶部
104d
統計量推定部
104e
第1近似部
標準モデル
122
112
書き込み部
103
参照モデル記憶部
参照モデル
121
102
参照モデル準備部
111
読み込み部

図2

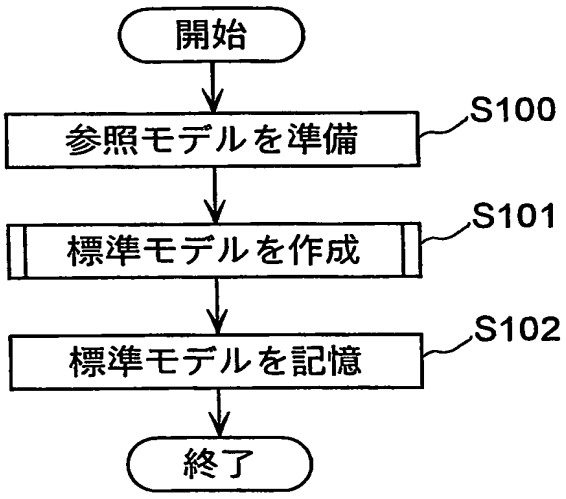


図3

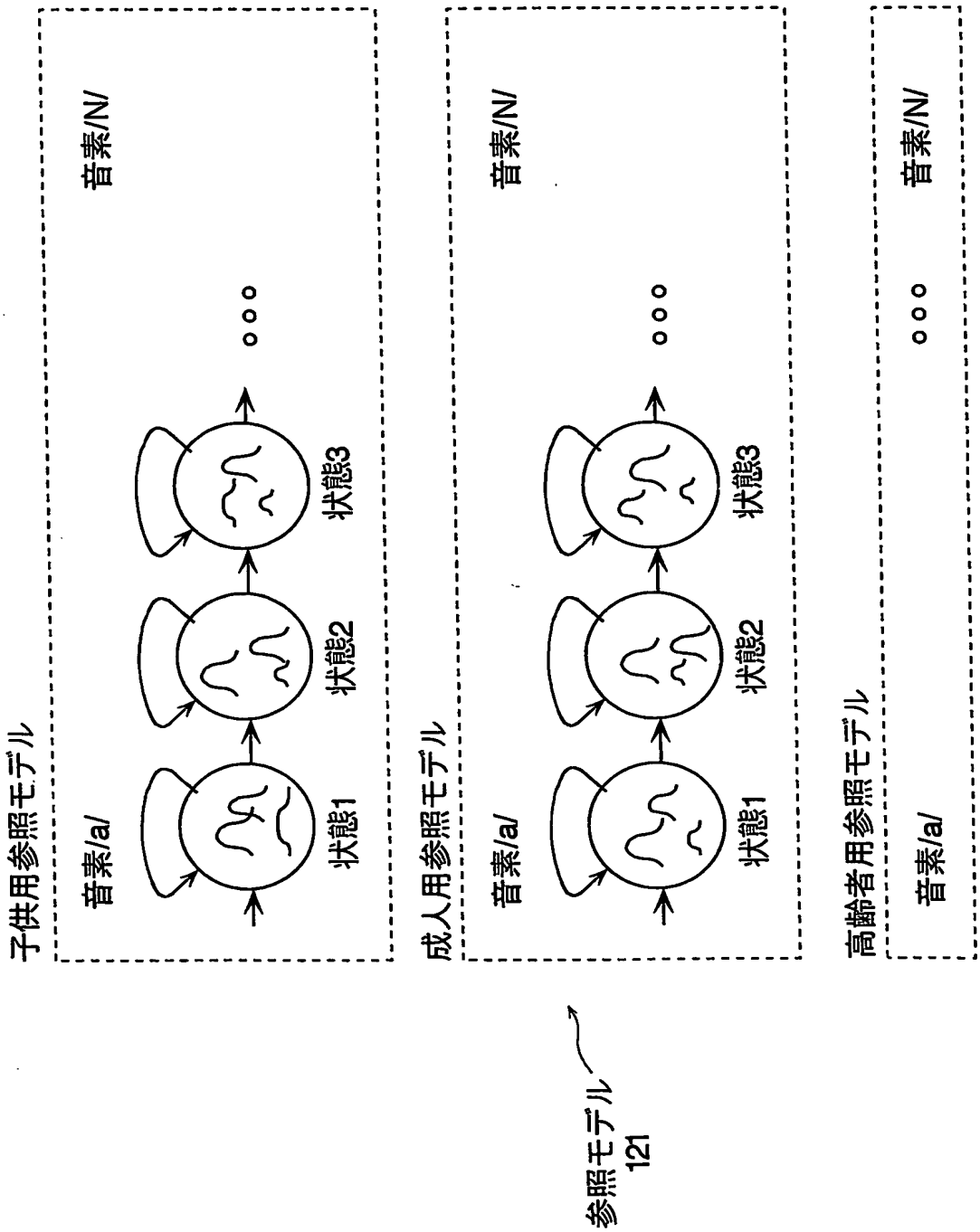

子供用参照モデル
音素/a/
状態1
状態2
状態3
ooo
音素/N/
成人用参照モデル
音素/a/
状態1
状態2
状態3
ooo
音素/N/
高齢者用参照モデル
音素/a/
ooo
音素/N/
参照モデル
121

図4

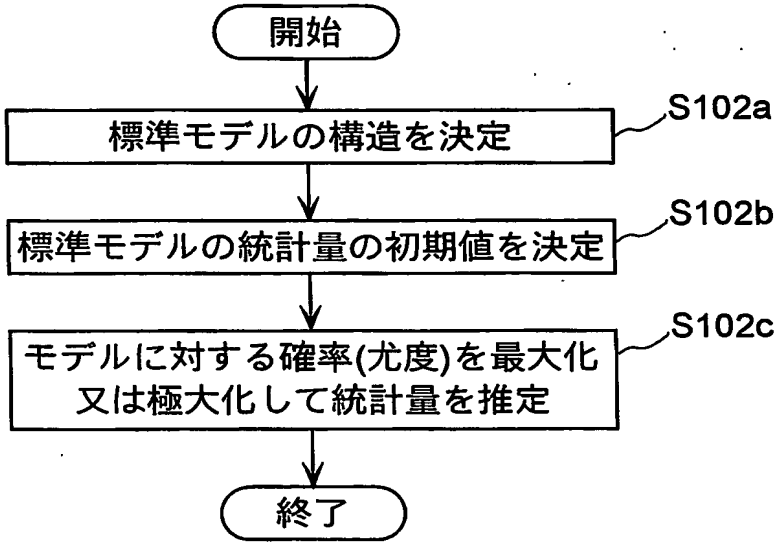

開始
標準モデルの構造を決定
S102a
標準モデルの統計量の初期値を決定
S102b
モデルに対する確率(尤度)を最大化
又は極大化して統計量を推定
S102c
終了

図5

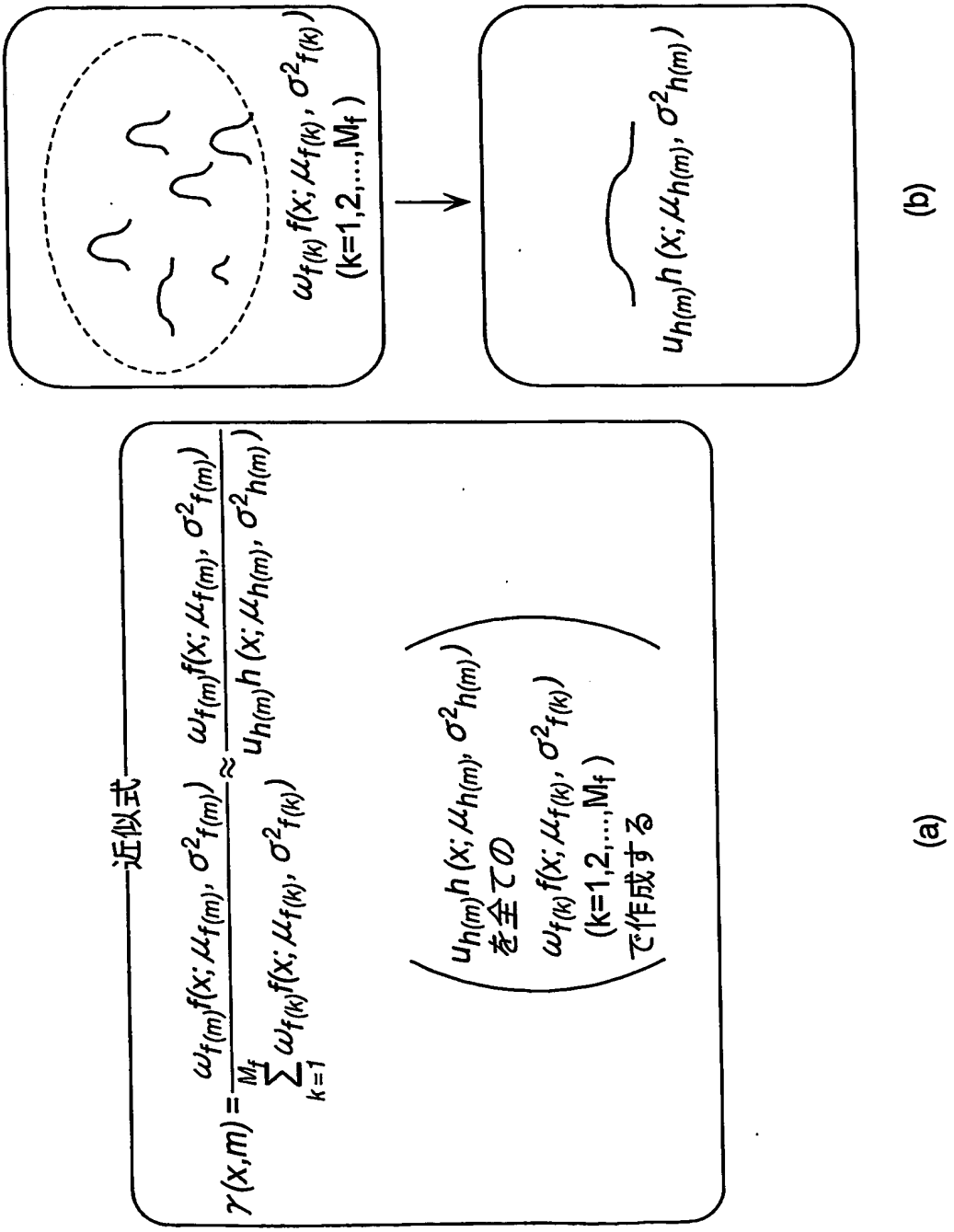
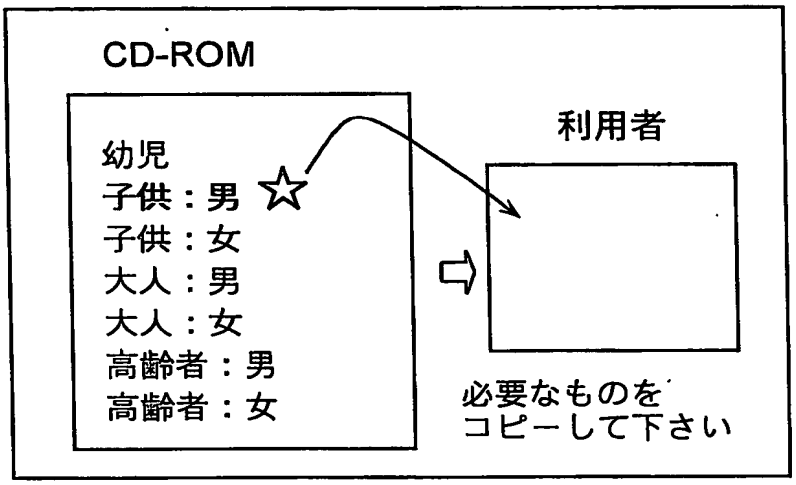

近似式
$\gamma(x,m) = \frac{\omega_{f(m)} f(x;\mu_{f(m)}, \sigma^2_{f(m)})}{\sum_{k=1}^{M_f} \omega_{f(k)} f(x;\mu_{f(k)}, \sigma^2_{f(k)})} \approx \frac{\omega_{f(m)} f(x;\mu_{f(m)}, \sigma^2_{f(m)})}{u_{h(m)} h(x;\mu_{h(m)}, \sigma^2_{h(m)})}$
$u_{h(m)} h(x;\mu_{h(m)}, \sigma^2_{h(m)})$
を全ての
$\omega_{f(k)} f(x;\mu_{f(k)}, \sigma^2_{f(k)})$
(k=1,2,...,M_f)
で作成する
(a)
$\omega_{f(k)} f(x;\mu_{f(k)}, \sigma^2_{f(k)})$
(k=1,2,...,M_f)
$u_{h(m)} h(x;\mu_{h(m)}, \sigma^2_{h(m)})$
(b)

図6

(a)


CD-ROM
幼児
子供：男
子供：女
大人：男
大人：女
高齢者：男
高齢者：女
利用者
必要なものを
コピーして下さい

(b)

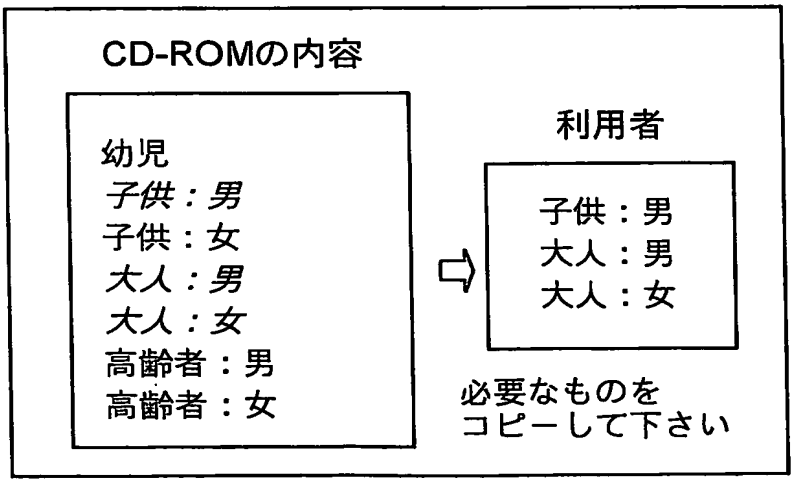
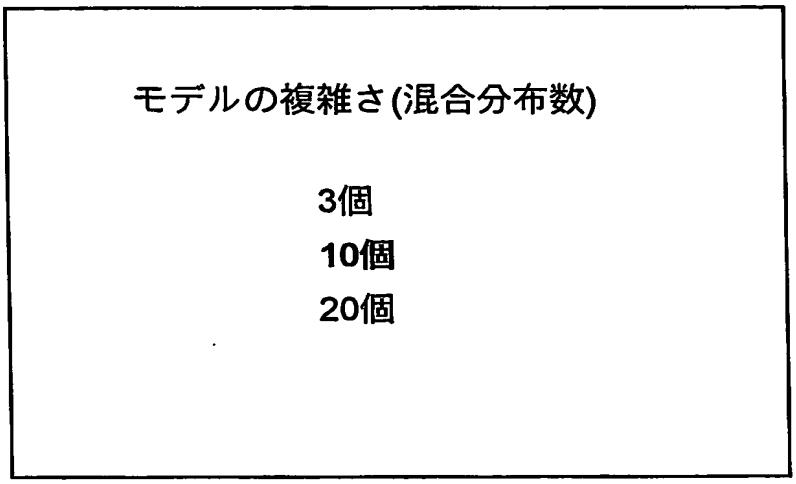

CD-ROMの内容
幼児
子供：男
子供：女
大人：男
大人：女
高齢者：男
高齢者：女
利用者
子供：男
大人：男
大人：女
必要なものを
コピーして下さい

図7

(a)



(b)

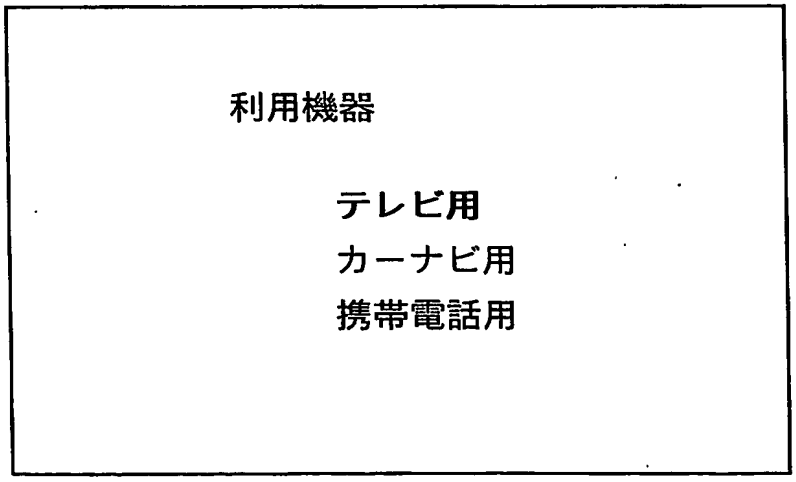
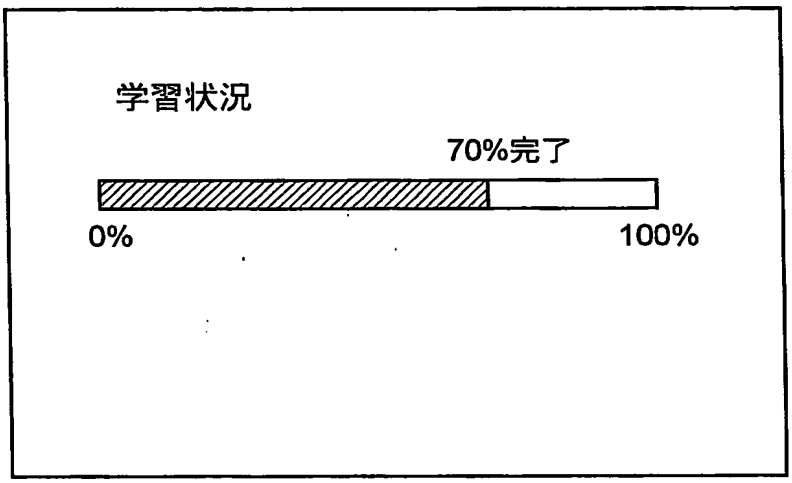


図8

(a)



(b)

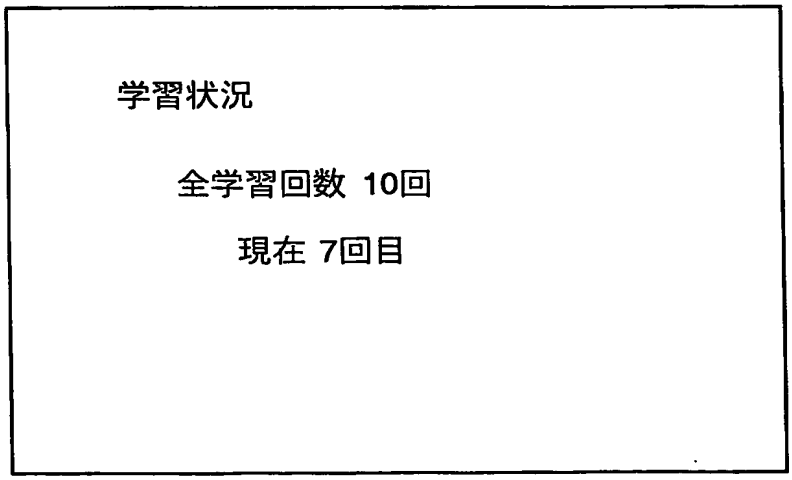


図9

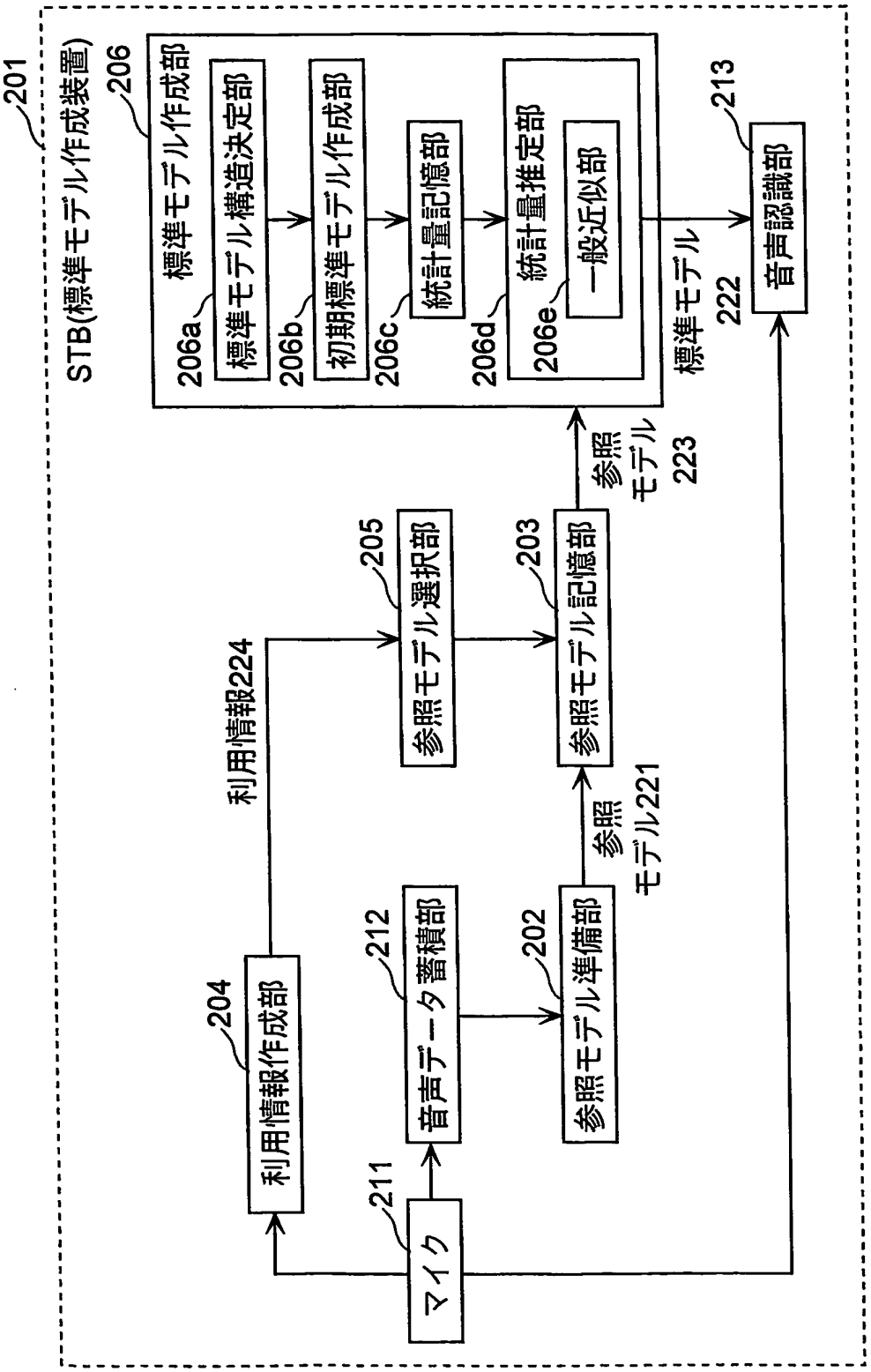

201
STB(標準モデル作成装置)
206
標準モデル作成部
206a
標準モデル構造決定部
206b
初期標準モデル作成部
206c
統計量記憶部
206d
統計量推定部
206e
一般近似部
標準モデル
222
213
音声認識部
参照モデル
223
205
参照モデル選択部
203
参照モデル記憶部
利用情報224
204
利用情報作成部
212
音声データ蓄積部
202
参照モデル準備部
参照モデル221
211
マイク

図10

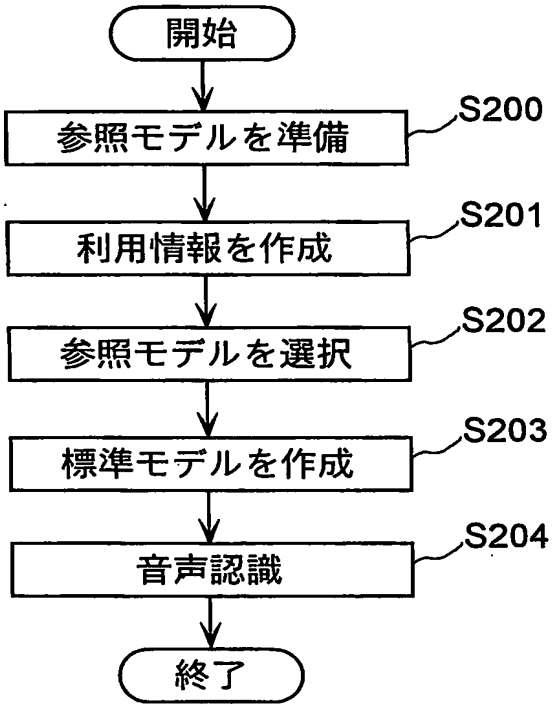

開始
参照モデルを準備
S200
利用情報を作成
S201
参照モデルを選択
S202
標準モデルを作成
S203
音声認識
S204
終了

図11

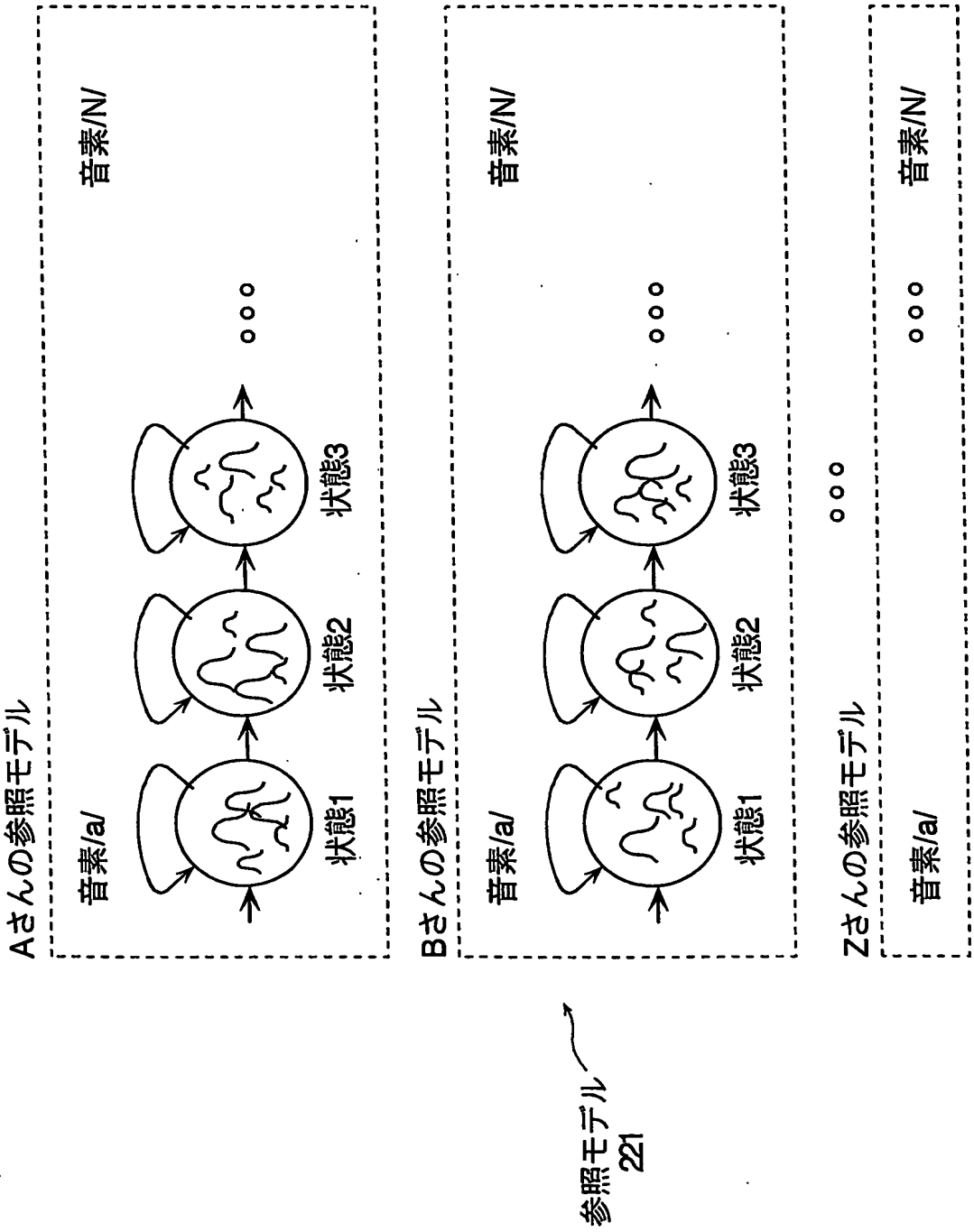

Aさんの参照モデル
音素/a/
状態1
状態2
状態3
音素/N/
Bさんの参照モデル
音素/a/
状態1
状態2
状態3
音素/N/
Zさんの参照モデル
音素/a/
音素/N/
参照モデル
221

図12

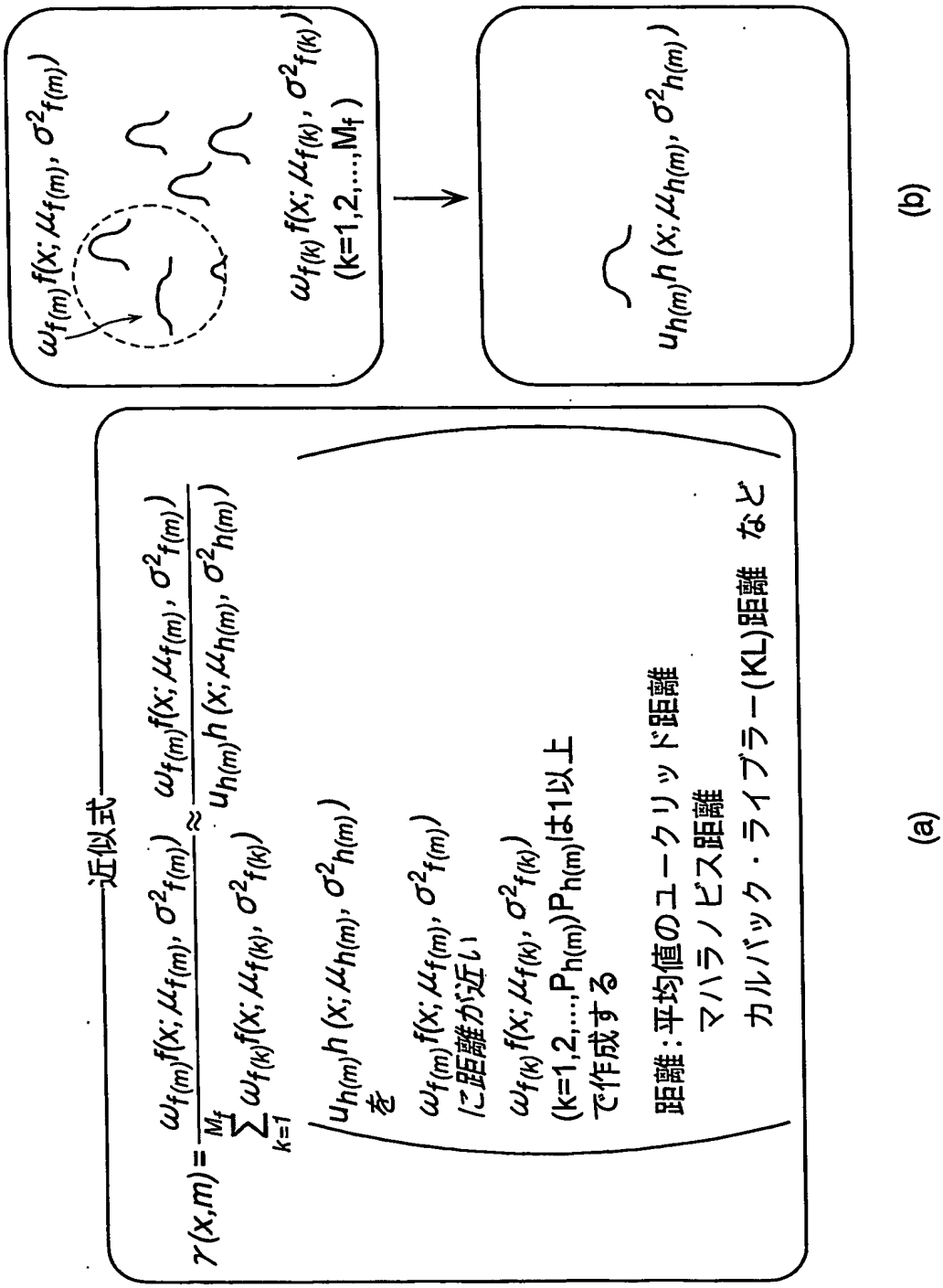

近似式
$\gamma(x,m)=\frac{\omega_{f(m)}f(x;\mu_{f(m)},\sigma^2_{f(m)})}{\sum_{k=1}^{M_f}\omega_{f(k)}f(x;\mu_{f(k)},\sigma^2_{f(k)})}\approx\frac{\omega_{f(m)}f(x;\mu_{f(m)},\sigma^2_{f(m)})}{u_{h(m)}h(x;\mu_{h(m)},\sigma^2_{h(m)})}$
$u_{h(m)}h(x;\mu_{h(m)},\sigma^2_{h(m)})$
を
$\omega_{f(m)}f(x;\mu_{f(m)},\sigma^2_{f(m)})$
に距離が近い
$\omega_{f(k)}f(x;\mu_{f(k)},\sigma^2_{f(k)})$
$(k=1,2,\ldots,P_{h(m)})P_{h(m)}$は1以上
で作成する
距離：平均値のユークリッド距離
マハラノビス距離
カルバック・ライブラー(KL)距離　など
(a)
$\omega_{f(m)}f(x;\mu_{f(m)},\sigma^2_{f(m)})$
$\omega_{f(k)}f(x;\mu_{f(k)},\sigma^2_{f(k)})$
$(k=1,2,\ldots,M_f)$
$u_{h(m)}h(x;\mu_{h(m)},\sigma^2_{h(m)})$
(b)

図13

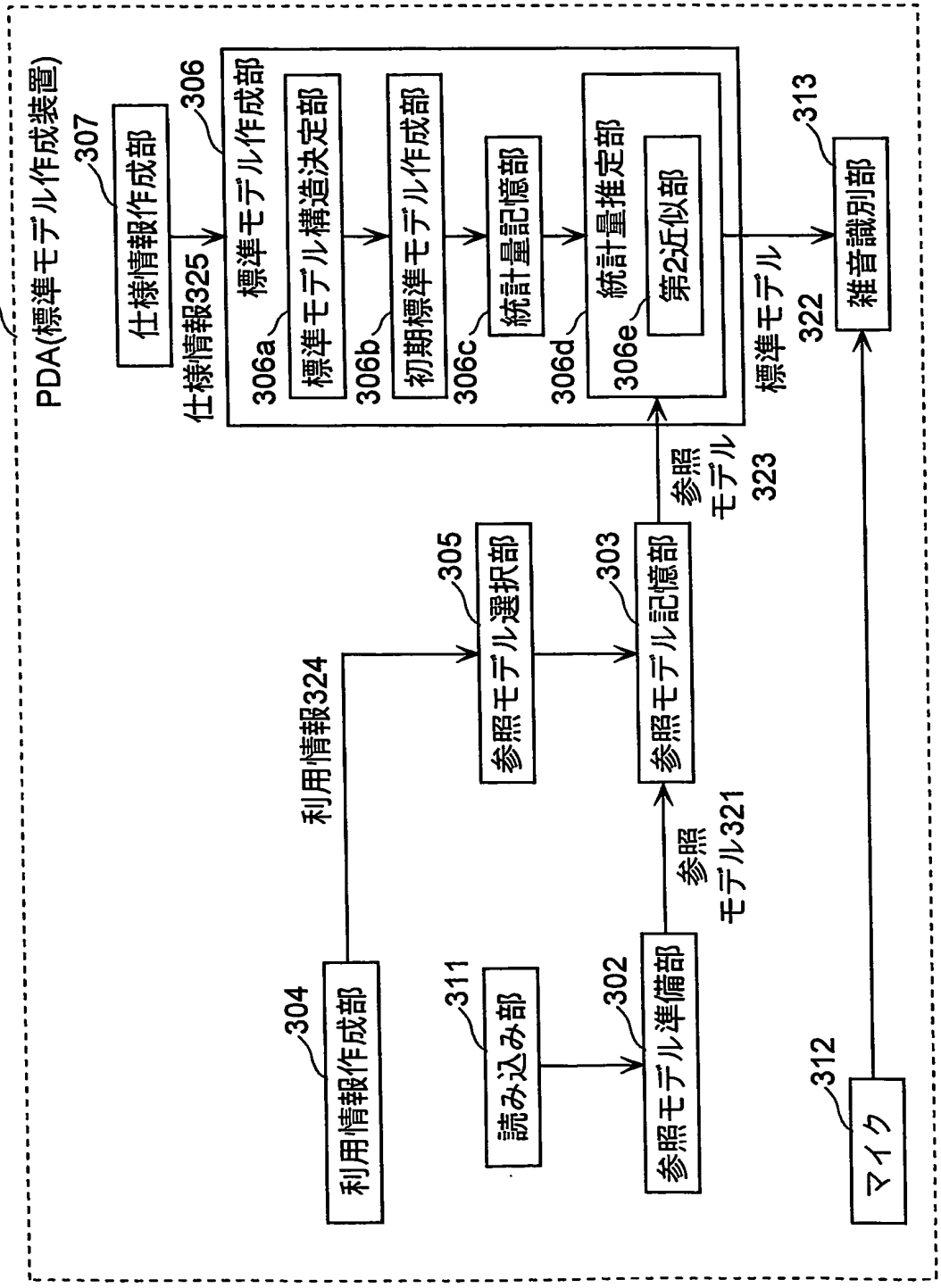

301
PDA(標準モデル作成装置)
307
仕様情報作成部
306
標準モデル作成部
仕様情報325
306a
標準モデル構造決定部
306b
初期標準モデル作成部
306c
統計量記憶部
306d
統計量推定部
306e
第2近似部
標準モデル
322
313
雑音識別部
参照
モデル
323
305
参照モデル選択部
303
参照モデル記憶部
利用情報324
参照
モデル321
304
利用情報作成部
311
読み込み部
302
参照モデル準備部
312
マイク

図14

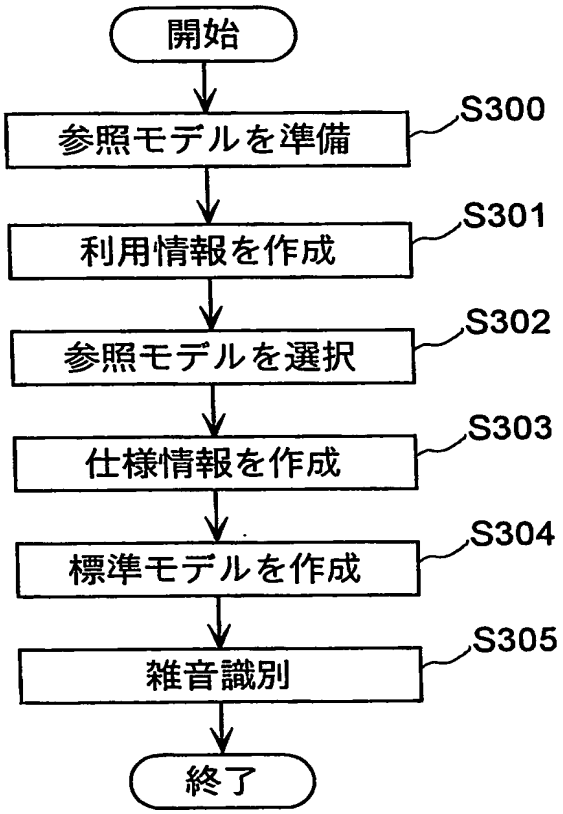

開始
参照モデルを準備
S300
利用情報を作成
S301
参照モデルを選択
S302
仕様情報を作成
S303
標準モデルを作成
S304
雑音識別
S305
終了

図15

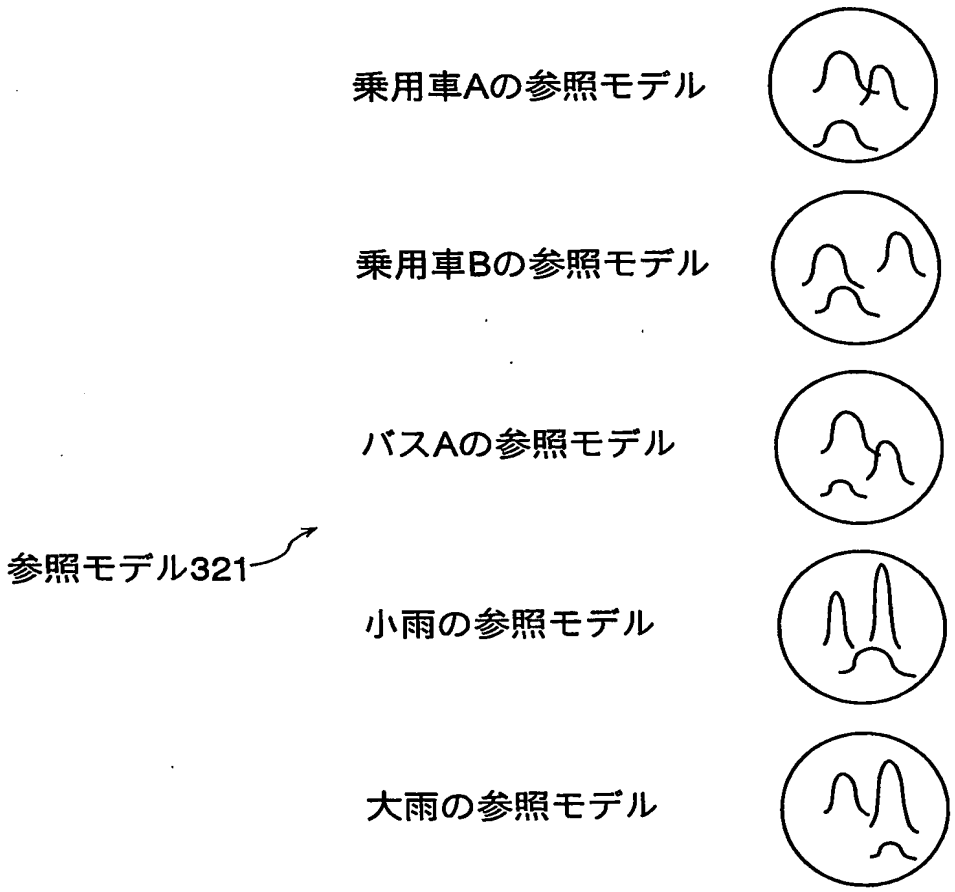

乗用車Aの参照モデル
乗用車Bの参照モデル
バスAの参照モデル
参照モデル321
小雨の参照モデル
大雨の参照モデル

図16

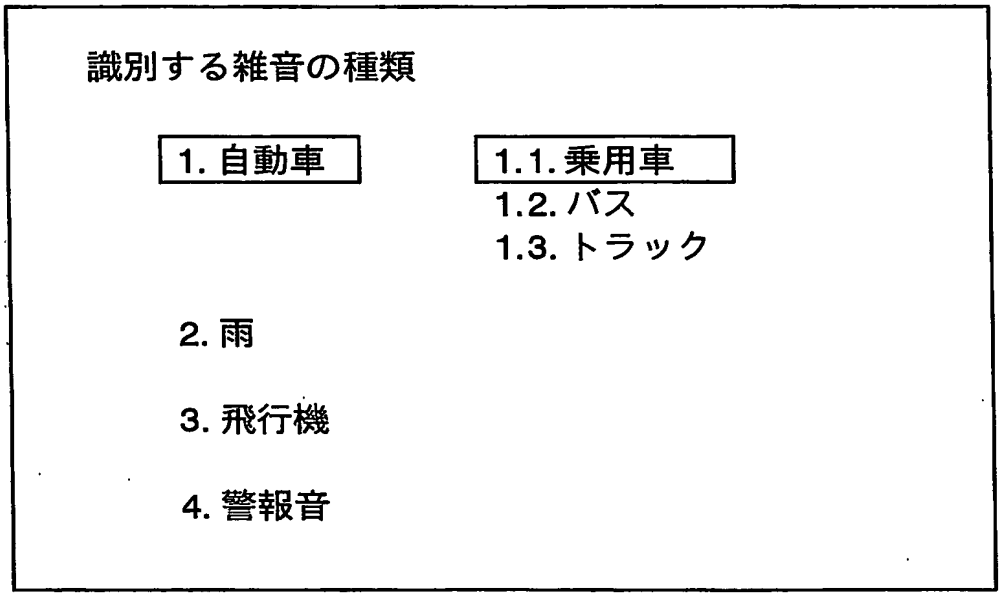


図17

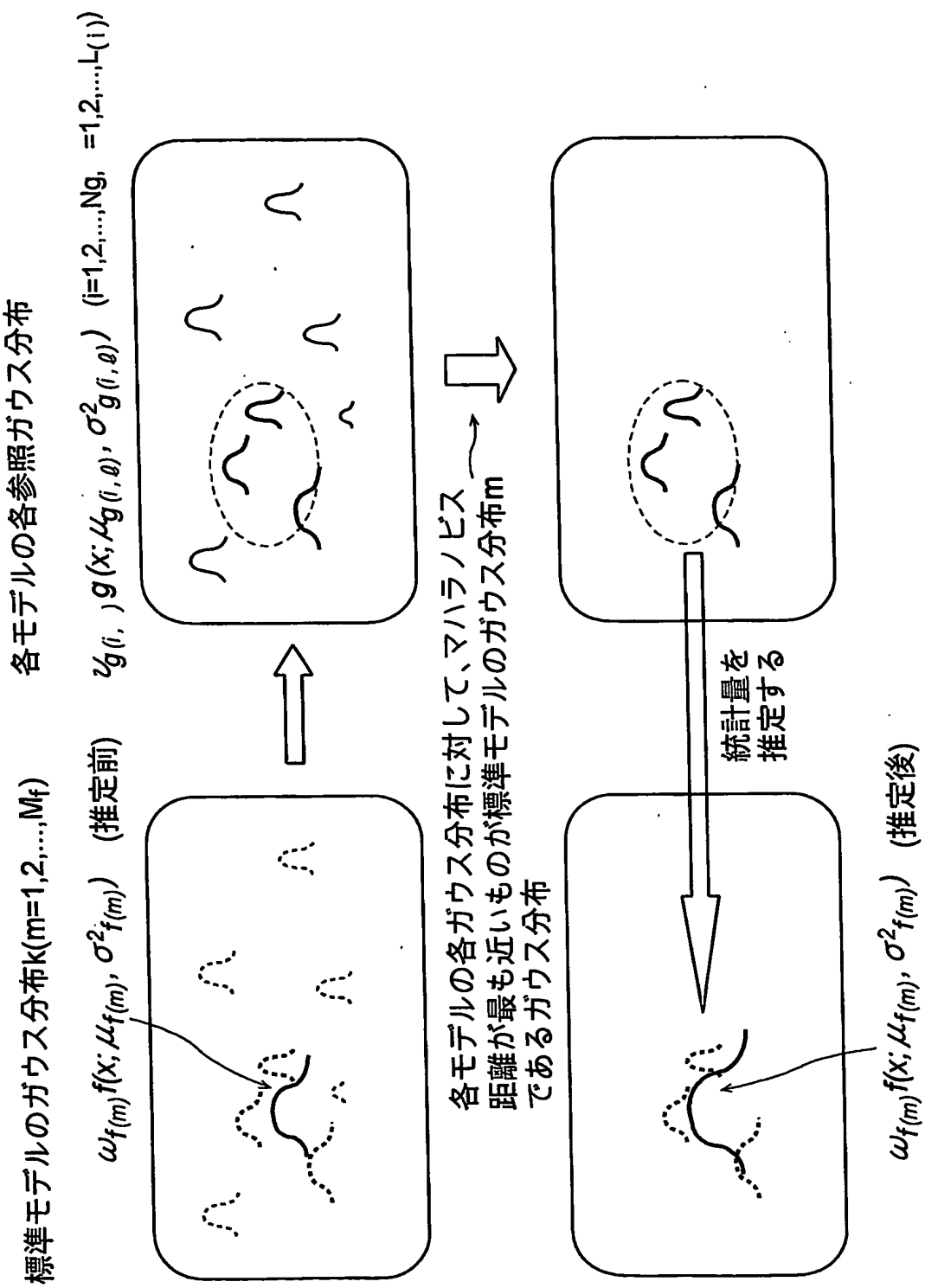

標準モデルのガウス分布k(m=1,2,...,Mf)
各モデルの各参照ガウス分布
(推定前)
(i=1,2,...,Ng, =1,2,...,L(i))
各モデルの各ガウス分布に対して、マハラノビス
距離が最も近いものが標準モデルのガウス分布m
であるガウス分布
統計量を
推定する
(推定後)

図18

近似式
$\gamma(x,m)=\frac{\omega_{f(m)}f(x;\mu_{f(m)},\sigma^2_{f(m)})}{\sum_{k=1}^{M_f}\omega_{f(k)}f(x;\mu_{f(k)},\sigma^2_{f(k)})}\approx 1.0$
$\upsilon_{g(i,\ell)}g(x;\mu_{g(i,\ell)},\sigma^2_{g(i,\ell)})$ (i=1,2,...,Ng, ℓ=1,2,...,L(i))
に距離が一番近い
$\omega_{f(m)}f(x;\mu_{f(m)},\sigma^2_{f(m)})$ (m=1,2,...,Mf)
のインデックス(m)を決定する
$\omega_{f(m)}f(x;\mu_{f(m)},\sigma^2_{f(m)})$ (m=1,2,...,Mf)
の統計量をインデックス(m)が付与された
$\upsilon_{g(i,\ell)}g(x;\mu_{g(i,\ell)},\sigma^2_{g(i,\ell)})$ (i=1,2,...,Ng, =1,2,...,L(i))
を用いて「十分統計量による方法(非特2)」と
同様な統計処理計算により求める
距離：平均値のユークリッド距離
マハラノビス距離
カルバック・ライブラー(KL)距離 など
(a)
$\omega_{f(m)}f(x;\mu_{f(m)},\sigma^2_{f(m)})$
$\upsilon_{g(i,\ell)}g(x;\mu_{g(i,\ell)},\sigma^2_{g(i,\ell)})$
$\omega_{f(m)}f(x;\mu_{f(m)},\sigma^2_{f(m)})$
(m=1,2,...,Mf)
$\omega_{f(m)}f(x;\mu_{f(m)},\sigma^2_{f(m)})$
(b)

図19

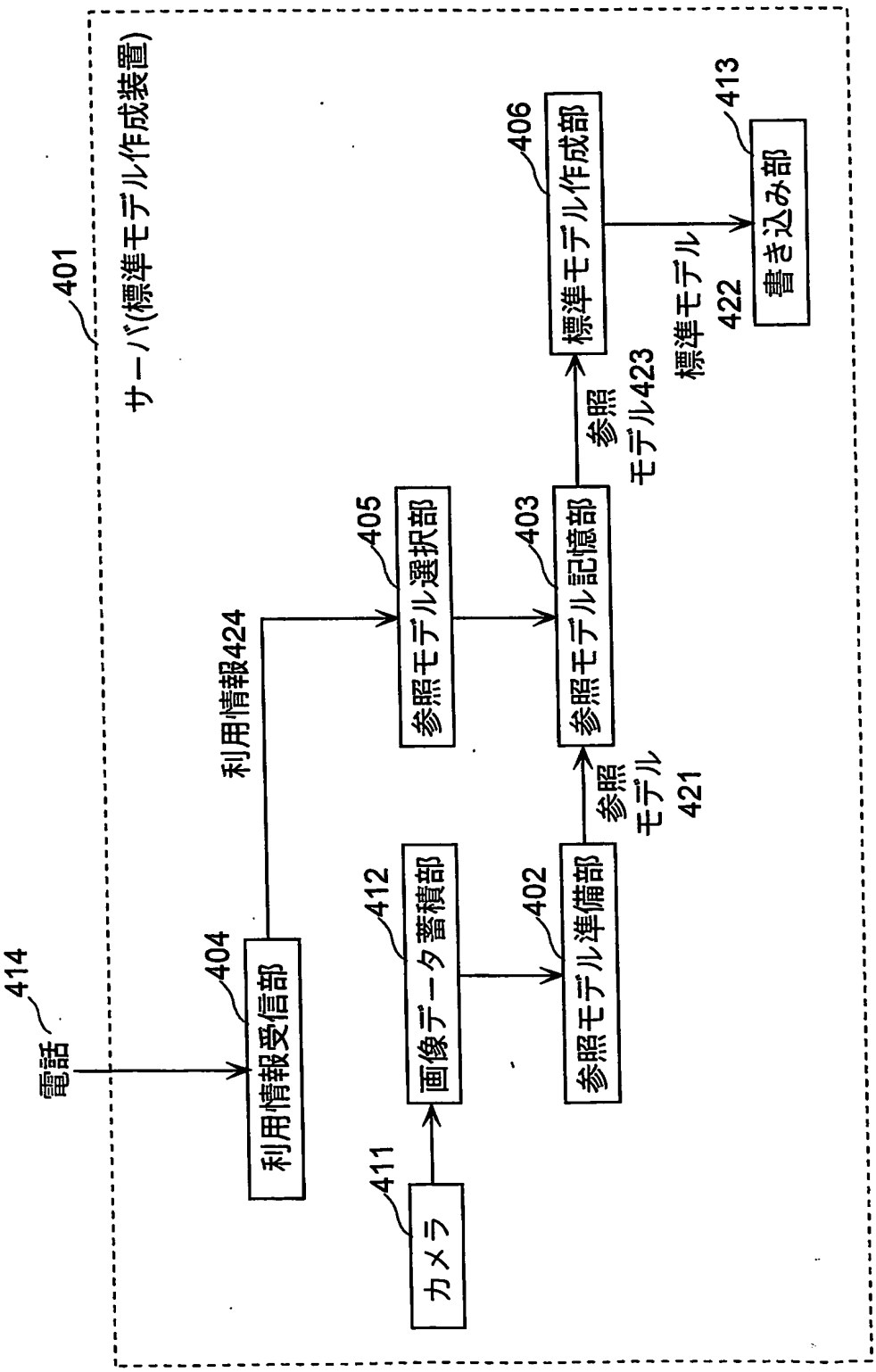

サーバ(標準モデル作成装置)
401
414
電話
404
利用情報受信部
利用情報424
405
参照モデル選択部
412
画像データ蓄積部
411
カメラ
403
参照モデル記憶部
402
参照モデル準備部
参照モデル421
参照モデル423
406
標準モデル作成部
標準モデル422
413
書き込み部

図20

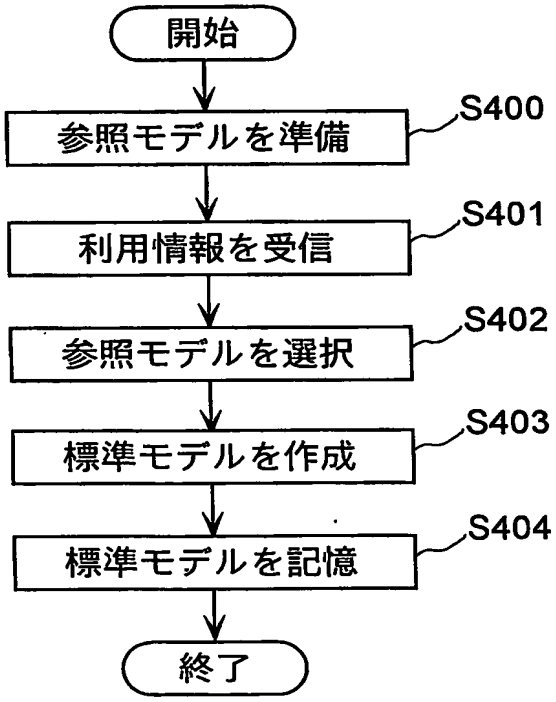

開始
参照モデルを準備
S400
利用情報を受信
S401
参照モデルを選択
S402
標準モデルを作成
S403
標準モデルを記憶
S404
終了

図21

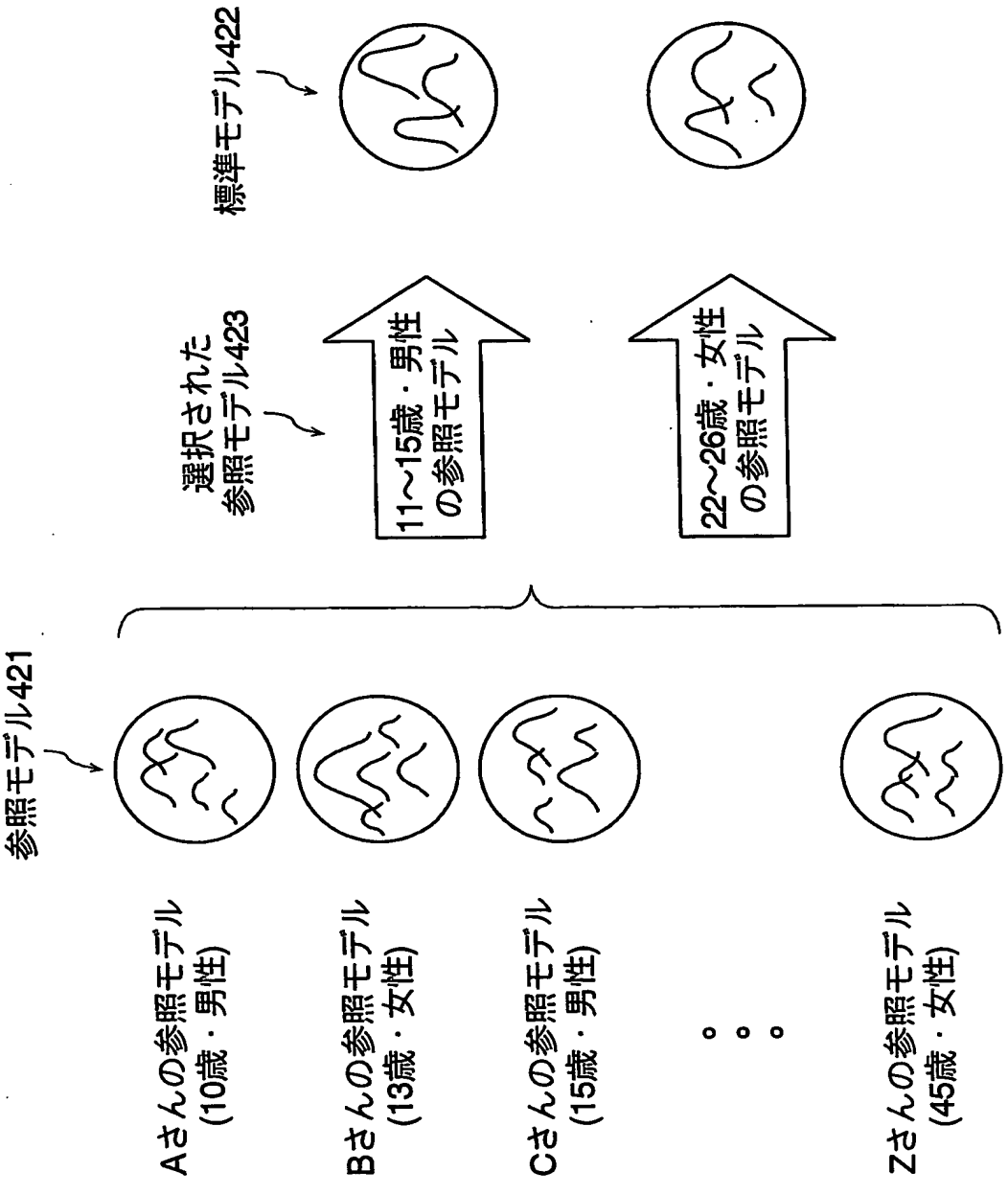

参照モデル421
選択された参照モデル423
標準モデル422
11～15歳・男性の参照モデル
22～26歳・女性の参照モデル
Aさんの参照モデル (10歳・男性)
Bさんの参照モデル (13歳・女性)
Cさんの参照モデル (15歳・男性)
Zさんの参照モデル (45歳・女性)

図22

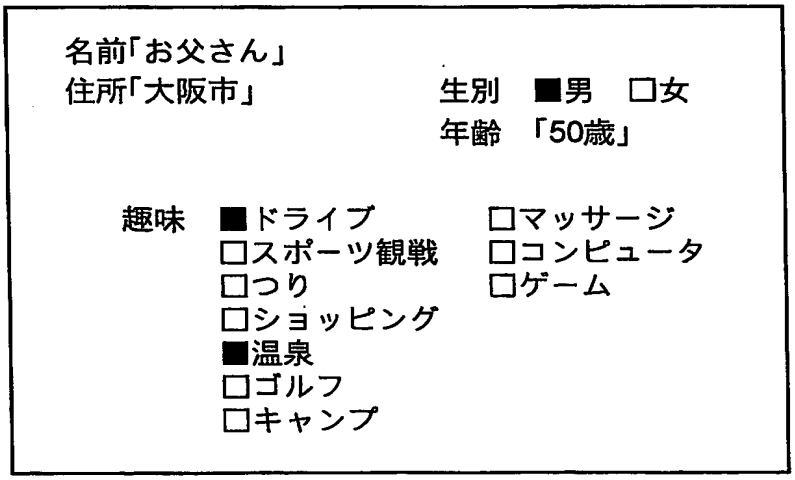

名前「お父さん」
住所「大阪市」
生別 ■男 □女
年齢 「50歳」
趣味 ■ドライブ
□スポーツ観戦
□つり
□ショッピング
■温泉
□ゴルフ
□キャンプ
□マッサージ
□コンピュータ
□ゲーム

図23

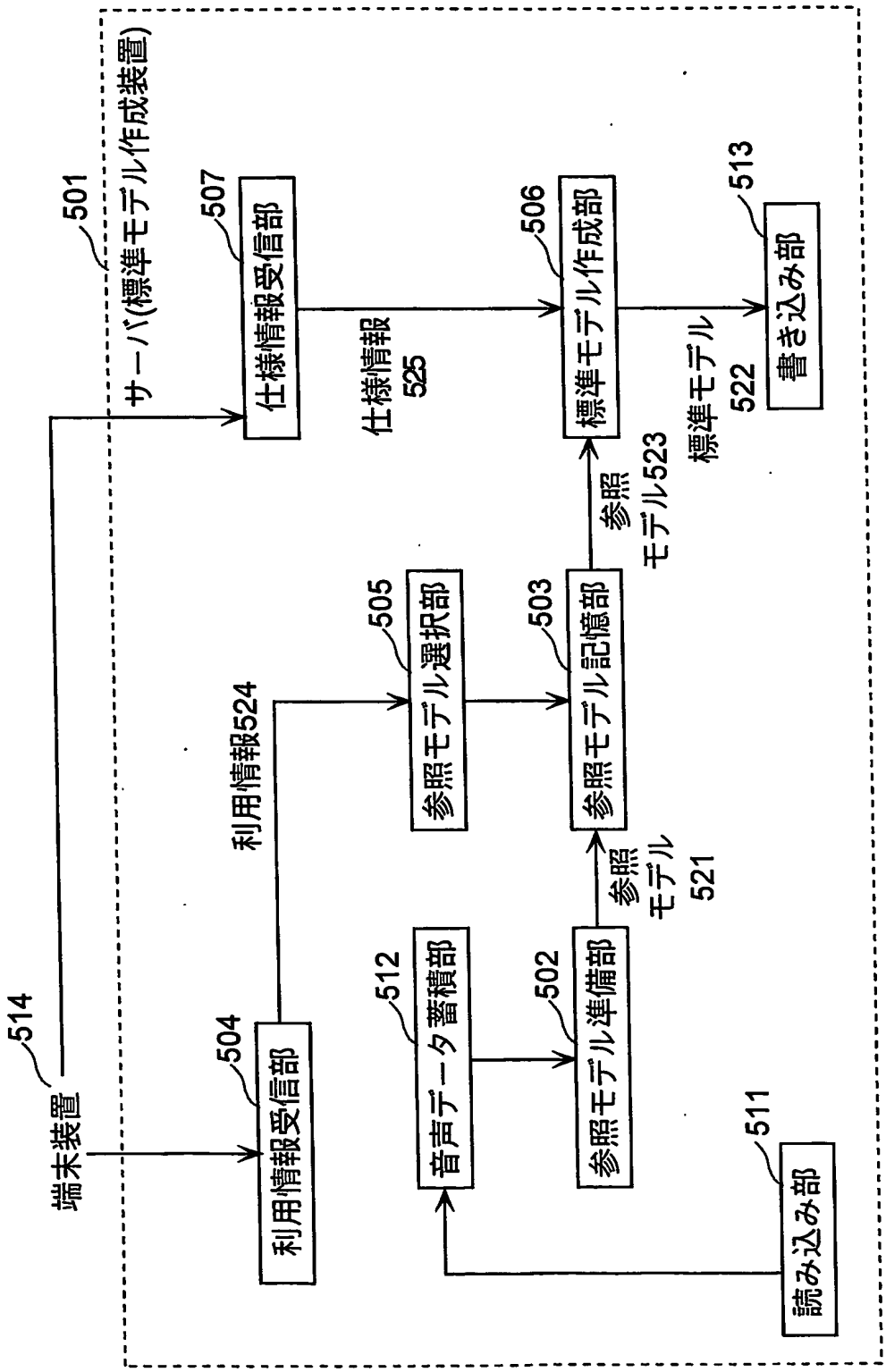

サーバ(標準モデル作成装置)
501
端末装置
514
仕様情報受信部
507
仕様情報
525
標準モデル作成部
506
書き込み部
513
標準モデル
522
参照
モデル523
利用情報524
利用情報受信部
504
参照モデル選択部
505
参照モデル記憶部
503
参照
モデル
521
音声データ蓄積部
512
参照モデル準備部
502
読み込み部
511

図24

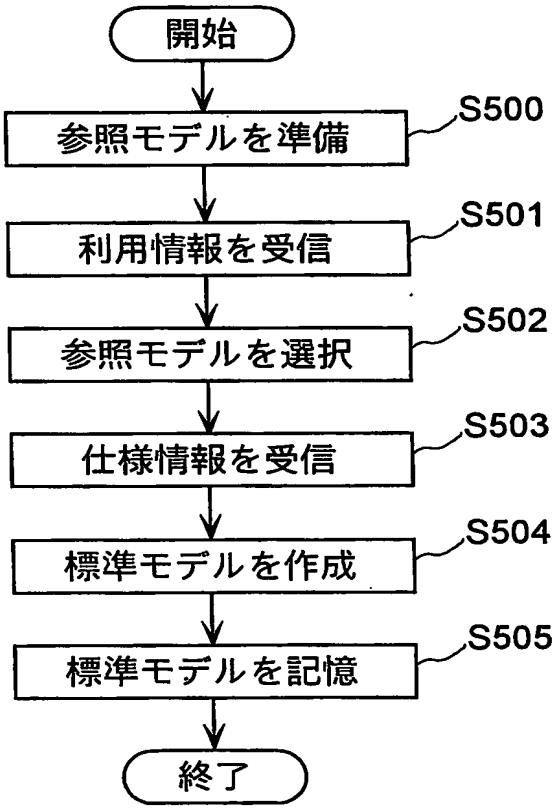

開始
参照モデルを準備
S500
利用情報を受信
S501
参照モデルを選択
S502
仕様情報を受信
S503
標準モデルを作成
S504
標準モデルを記憶
S505
終了

図25

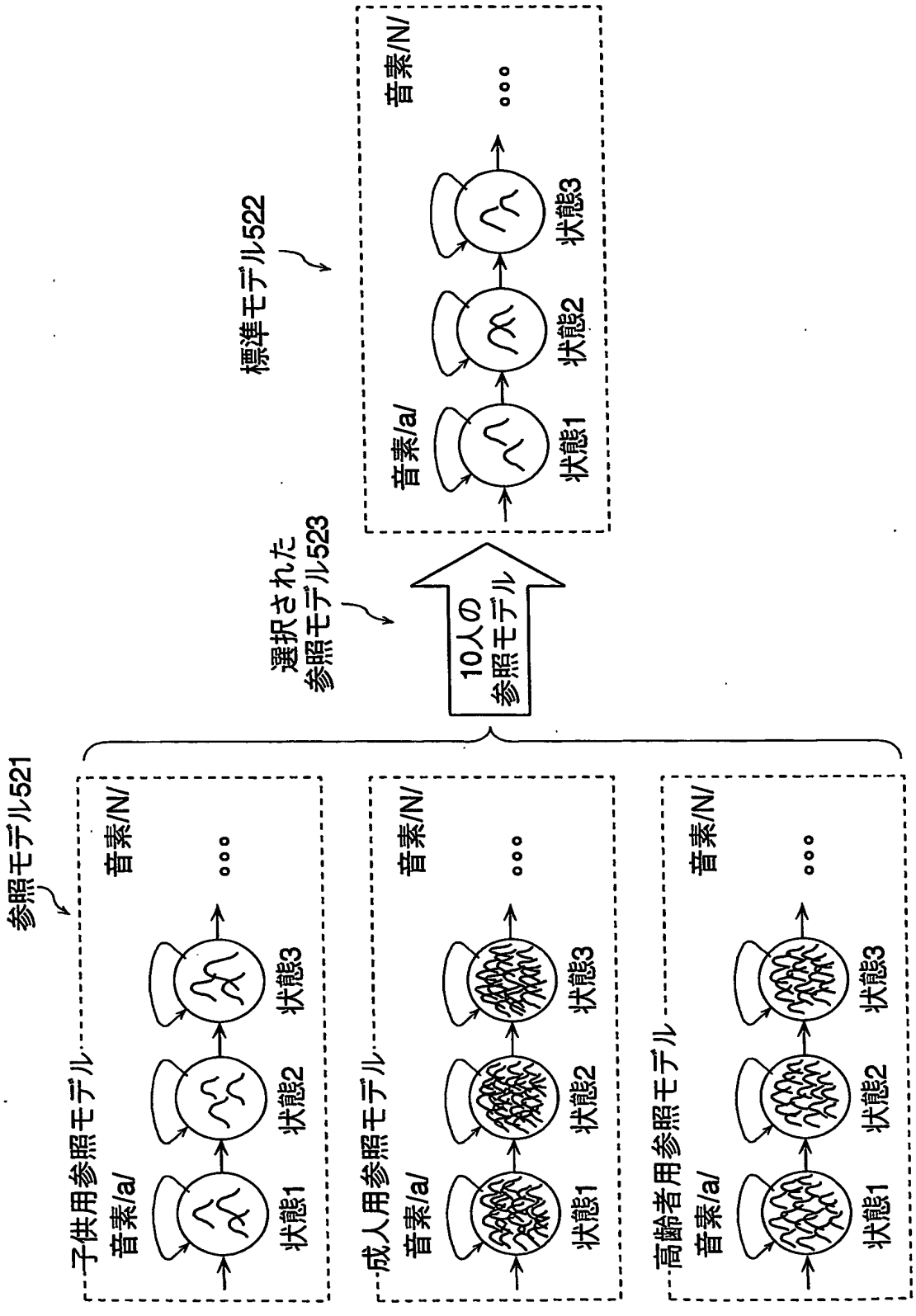

参照モデル521
子供用参照モデル
音素/a/
状態1
状態2
状態3
音素/N/
成人用参照モデル
高齢者用参照モデル
10人の
参照モデル
選択された
参照モデル523
標準モデル522

図26

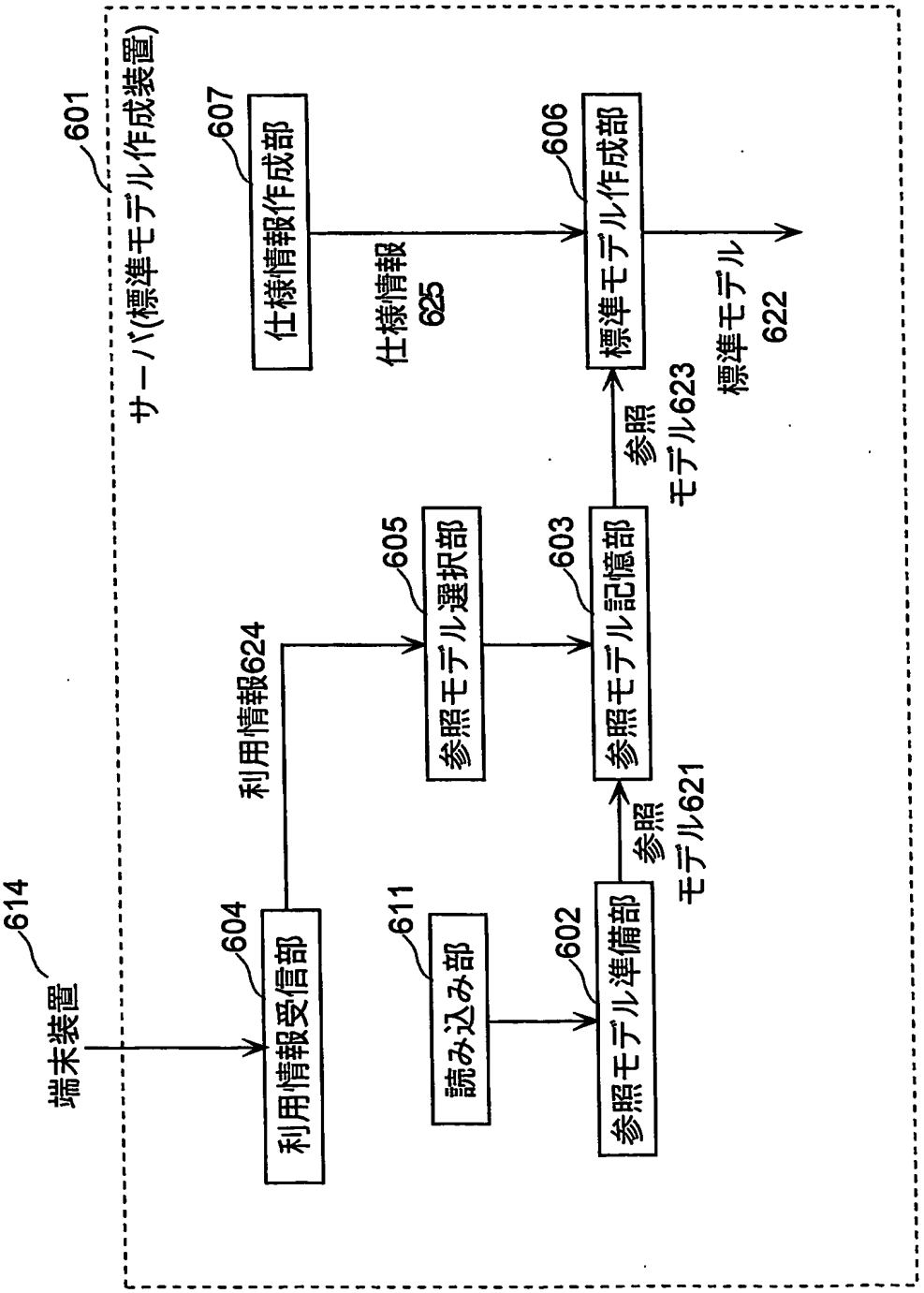

601
サーバ(標準モデル作成装置)
607
仕様情報作成部
606
標準モデル作成部
仕様情報
625
標準モデル
622
参照
モデル623
605
参照モデル選択部
603
参照モデル記憶部
利用情報624
614
端末装置
604
利用情報受信部
611
読み込み部
602
参照モデル準備部
参照
モデル621

図27

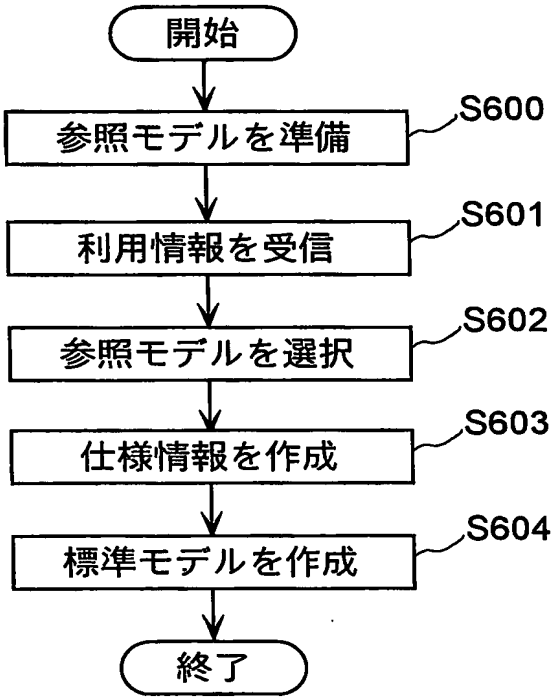

開始
参照モデルを準備
S600
利用情報を受信
S601
参照モデルを選択
S602
仕様情報を作成
S603
標準モデルを作成
S604
終了

図28

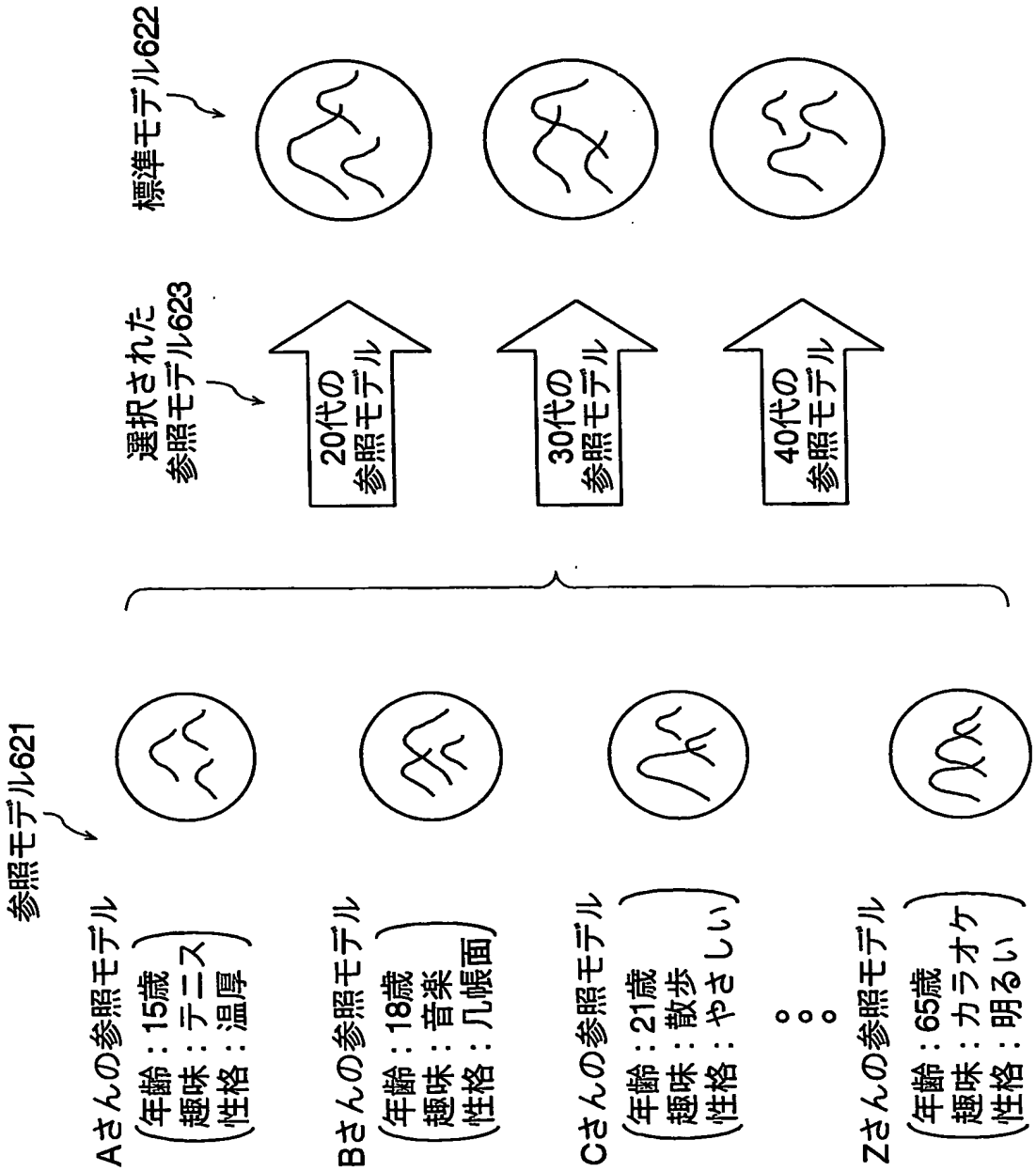

参照モデル621
Aさんの参照モデル
(年齢：15歳
趣味：テニス
性格：温厚)
Bさんの参照モデル
(年齢：18歳
趣味：音楽
性格：几帳面)
Cさんの参照モデル
(年齢：21歳
趣味：散歩
性格：やさしい)
。。。
Zさんの参照モデル
(年齢：65歳
趣味：カラオケ
性格：明るい)
選択された
参照モデル623
20代の
参照モデル
30代の
参照モデル
40代の
参照モデル
標準モデル622

図29

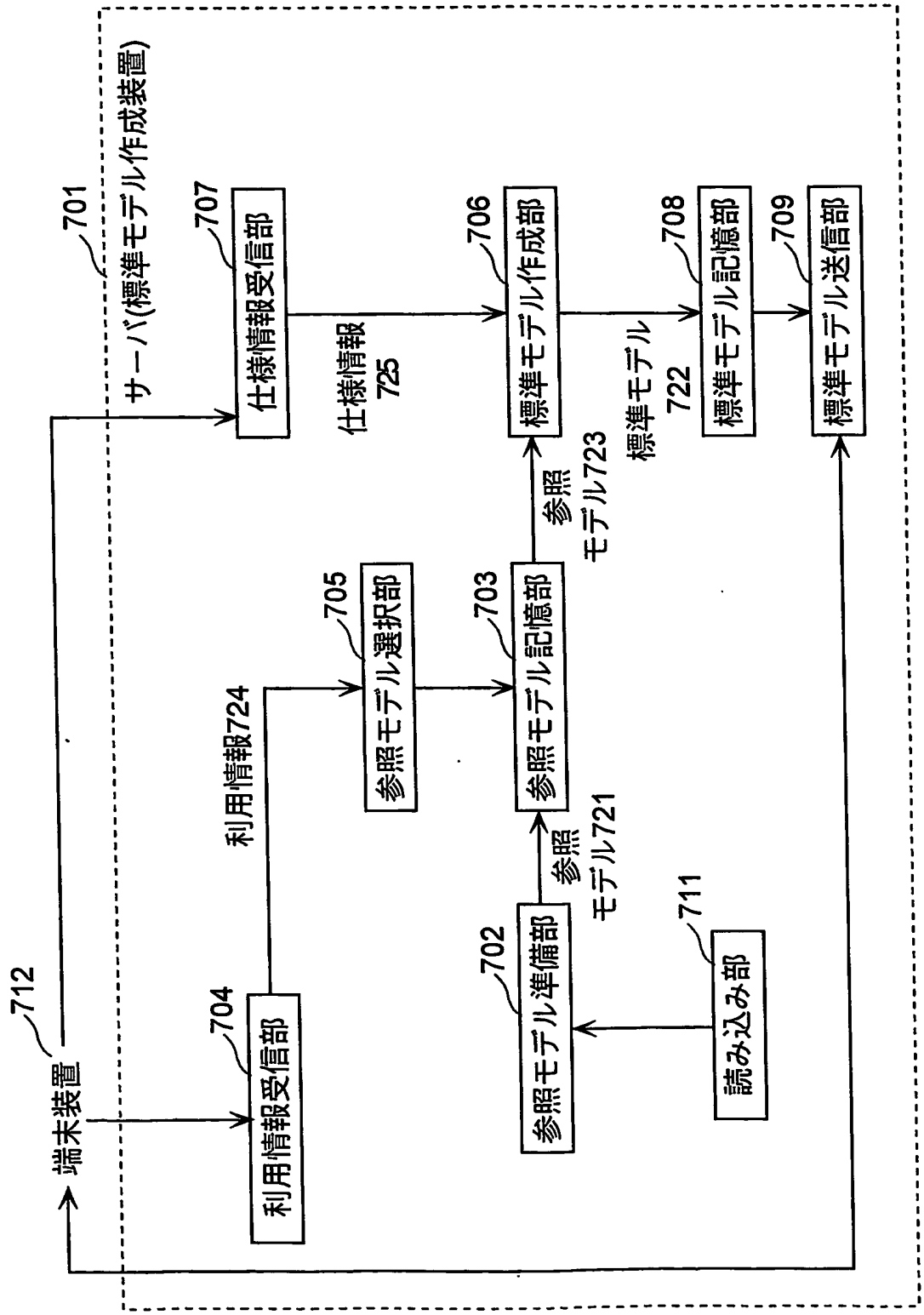

701
サーバ(標準モデル作成装置)
707
仕様情報受信部
仕様情報
725
706
標準モデル作成部
708
標準モデル記憶部
709
標準モデル送信部
標準モデル
722
参照
モデル723
705
参照モデル選択部
703
参照モデル記憶部
利用情報724
704
利用情報受信部
712
端末装置
702
参照モデル準備部
参照
モデル721
711
読み込み部

図30

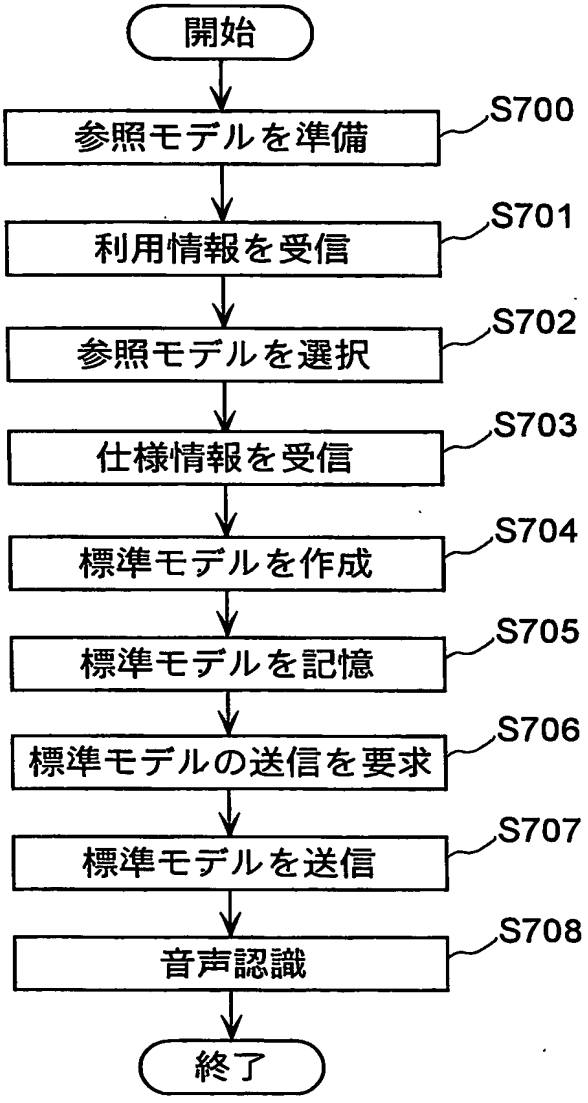

開始
参照モデルを準備
S700
利用情報を受信
S701
参照モデルを選択
S702
仕様情報を受信
S703
標準モデルを作成
S704
標準モデルを記憶
S705
標準モデルの送信を要求
S706
標準モデルを送信
S707
音声認識
S708
終了

図31

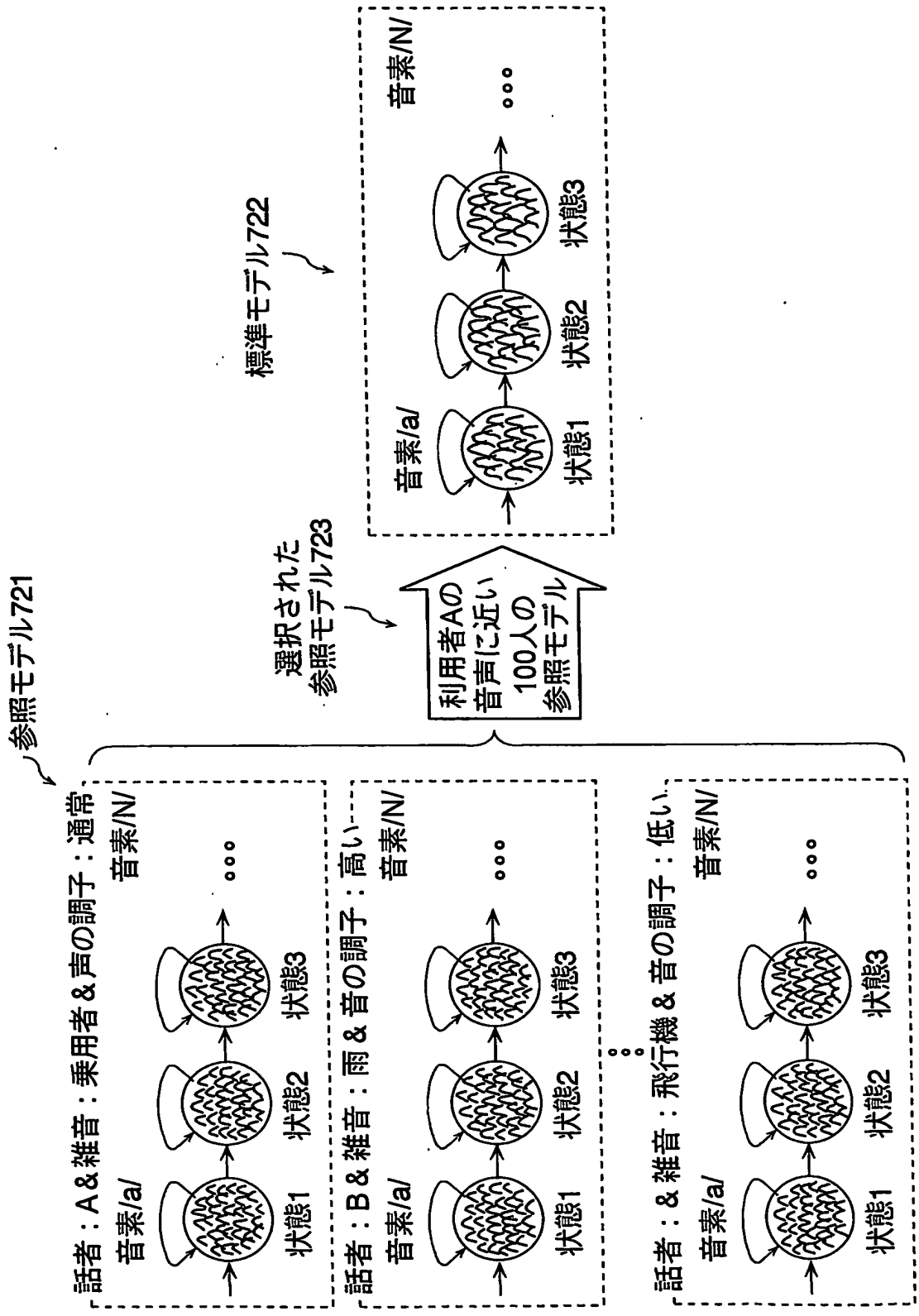

参照モデル721
話者：A＆雑音：乗用者＆声の調子：通常
音素/N/
音素/a/
状態1
状態2
状態3
話者：B＆雑音：雨＆音の調子：高い
話者：＆雑音：飛行機＆音の調子：低い
選択された
参照モデル723
利用者Aの
音声に近い
100人の
参照モデル
標準モデル722

図32

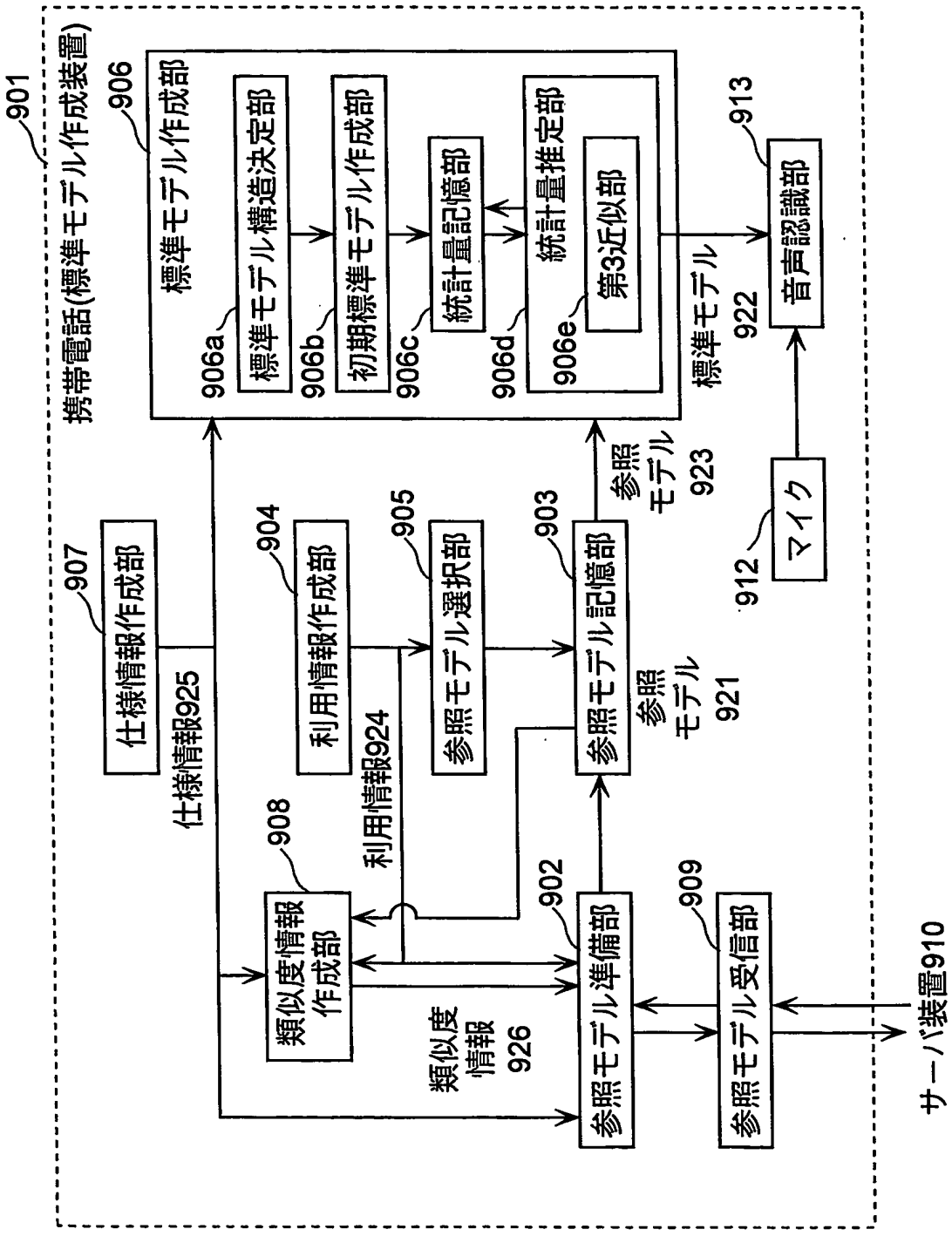

901
携帯電話(標準モデル作成装置)
906
標準モデル作成部
906a
標準モデル構造決定部
906b
初期標準モデル作成部
906c
統計量記憶部
906d
統計量推定部
906e
第3近似部
913
音声認識部
922
標準モデル
907
仕様情報作成部
仕様情報925
904
利用情報作成部
利用情報924
905
参照モデル選択部
903
参照モデル記憶部
参照モデル
923
参照モデル
921
912
マイク
908
類似度情報作成部
902
参照モデル準備部
909
参照モデル受信部
類似度情報
926
サーバ装置910

図33

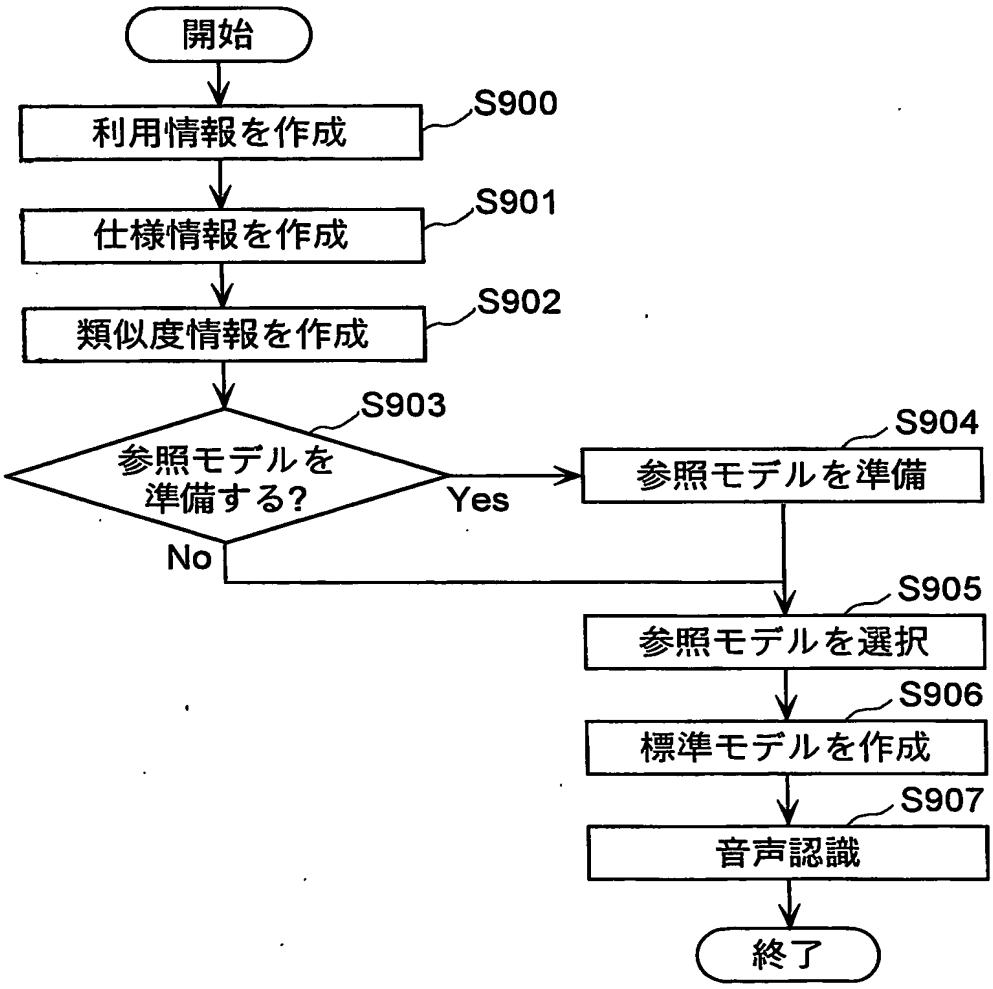

開始
利用情報を作成
S900
仕様情報を作成
S901
類似度情報を作成
S902
参照モデルを
準備する?
S903
Yes
No
参照モデルを準備
S904
参照モデルを選択
S905
標準モデルを作成
S906
音声認識
S907
終了

図34

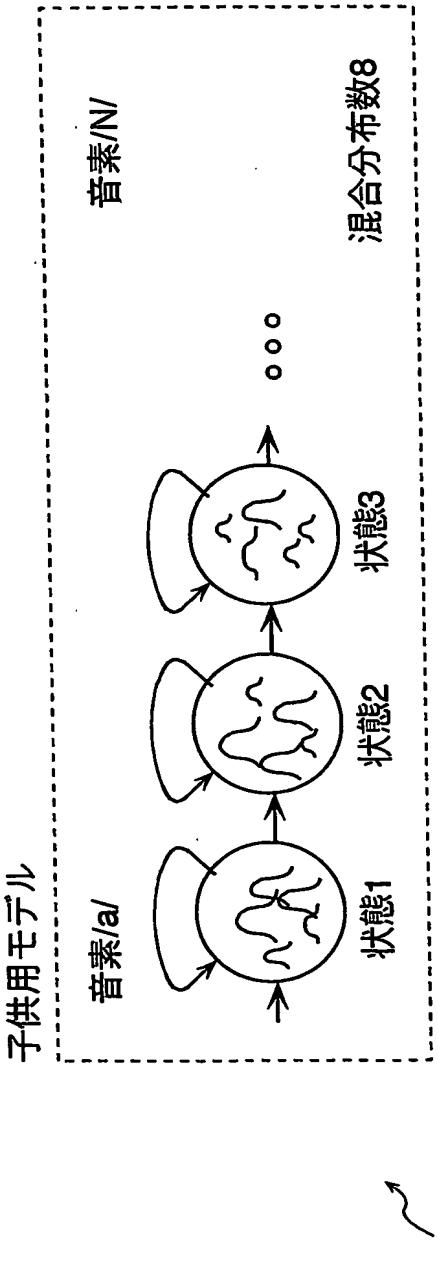


図35

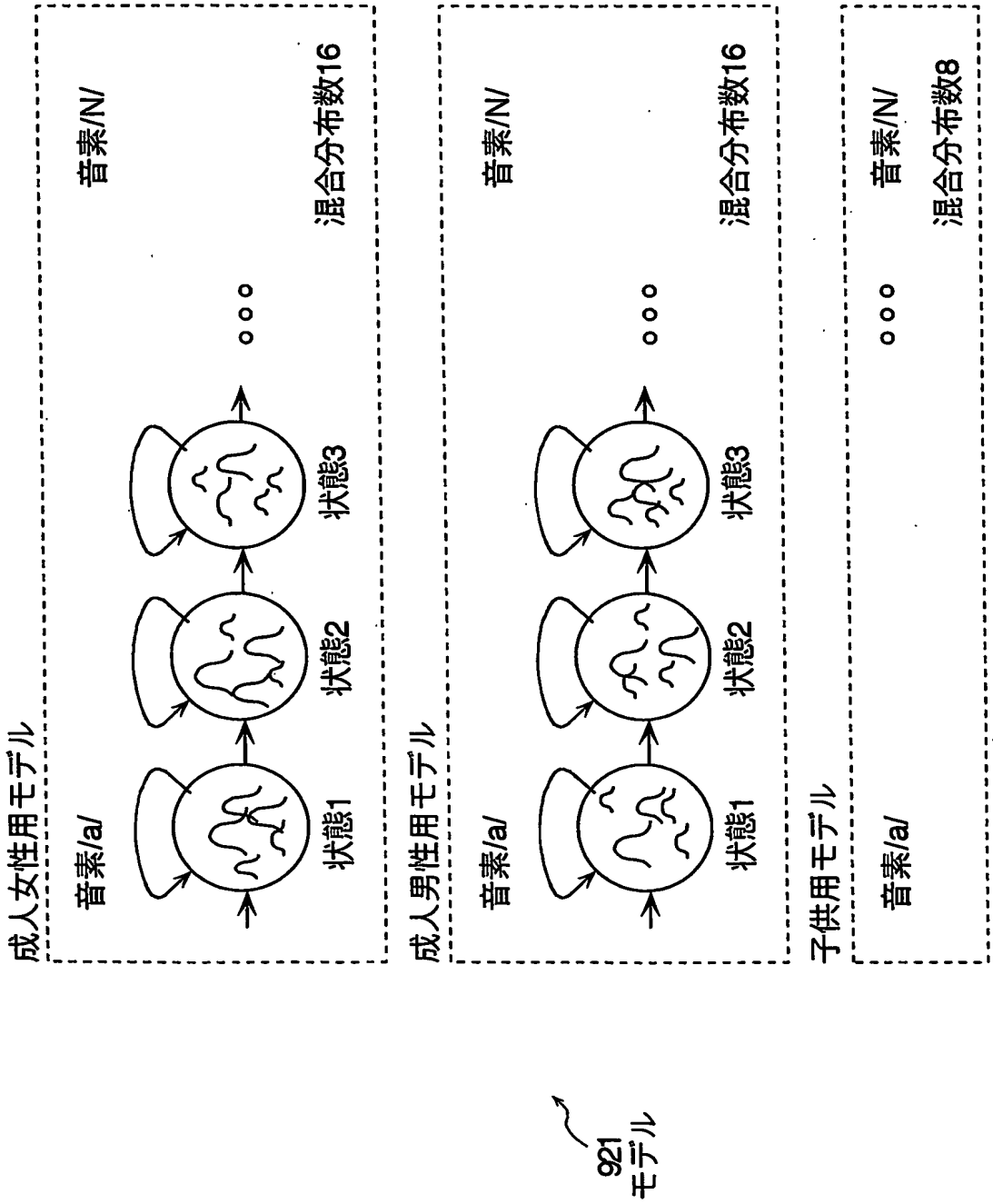
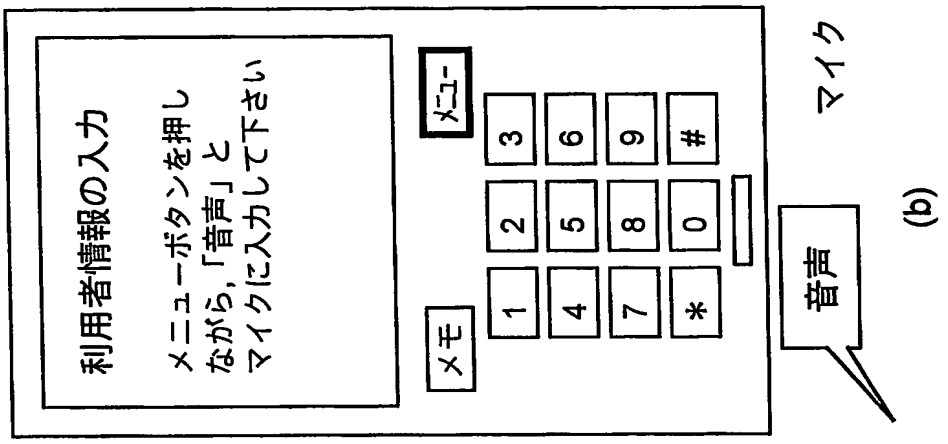

成人女性用モデル
音素/a/
音素/N/
状態1
状態2
状態3
混合分布数16
成人男性用モデル
音素/a/
音素/N/
状態1
状態2
状態3
混合分布数16
子供用モデル
音素/a/
音素/N/
混合分布数8
921
モデル

図36


利用者情報の入力
メニューボタンを押し
ながら,「音声」と
マイクに入力して下さい
メニュー
メモ
1
2
3
4
5
6
7
8
9
*
0
#
音声
マイク
(b)

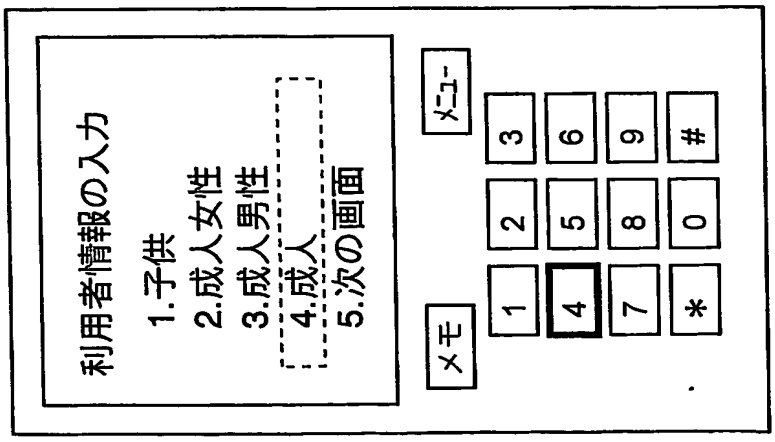
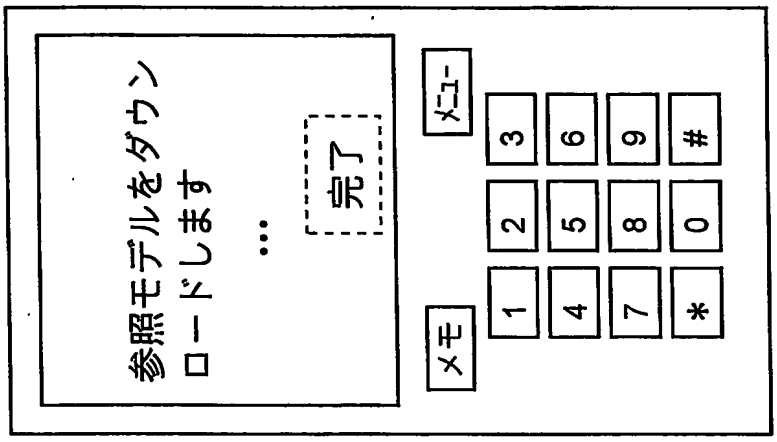
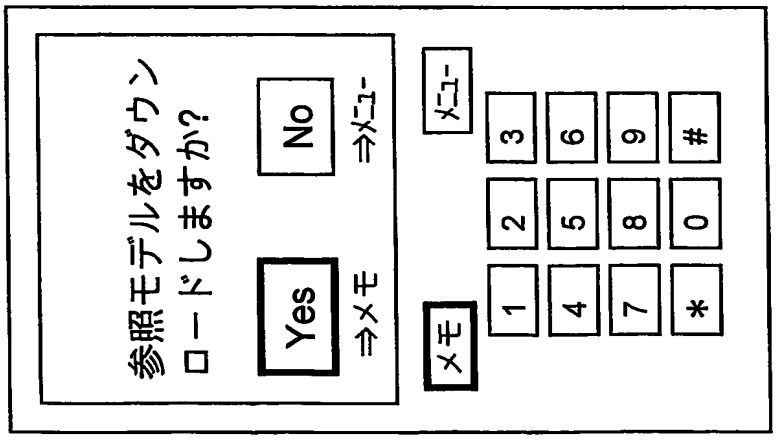

利用者情報の入力
1.子供
2.成人女性
3.成人男性
4.成人
5.次の画面
メニュー
メモ
1
2
3
4
5
6
7
8
9
*
0
#
(a)

図37


参照モデルをダウン
ロードします
...
完了
メニュー
メモ
1
2
3
4
5
6
7
8
9
*
0
#

(b)


参照モデルをダウン
ロードしますか?
Yes
⇒メモ
No
⇒メニュー
メニュー
メモ
1
2
3
4
5
6
7
8
9
*
0
#

(a)

図38

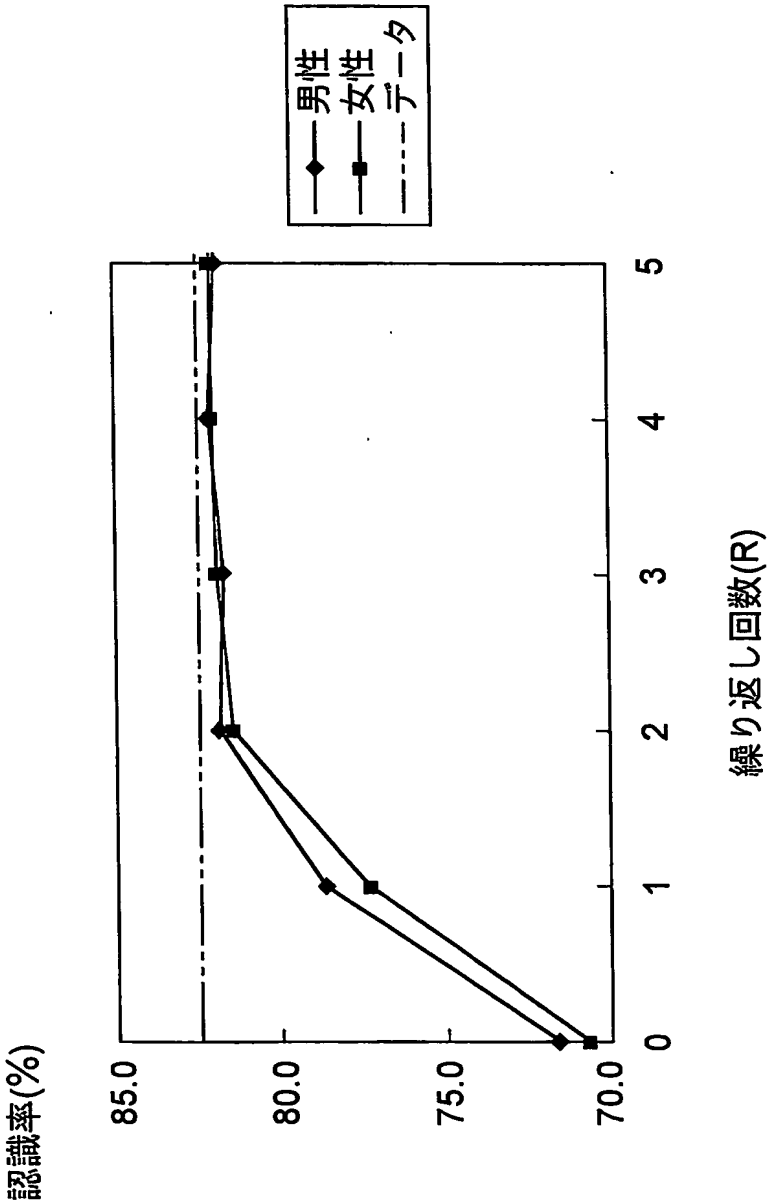


図39

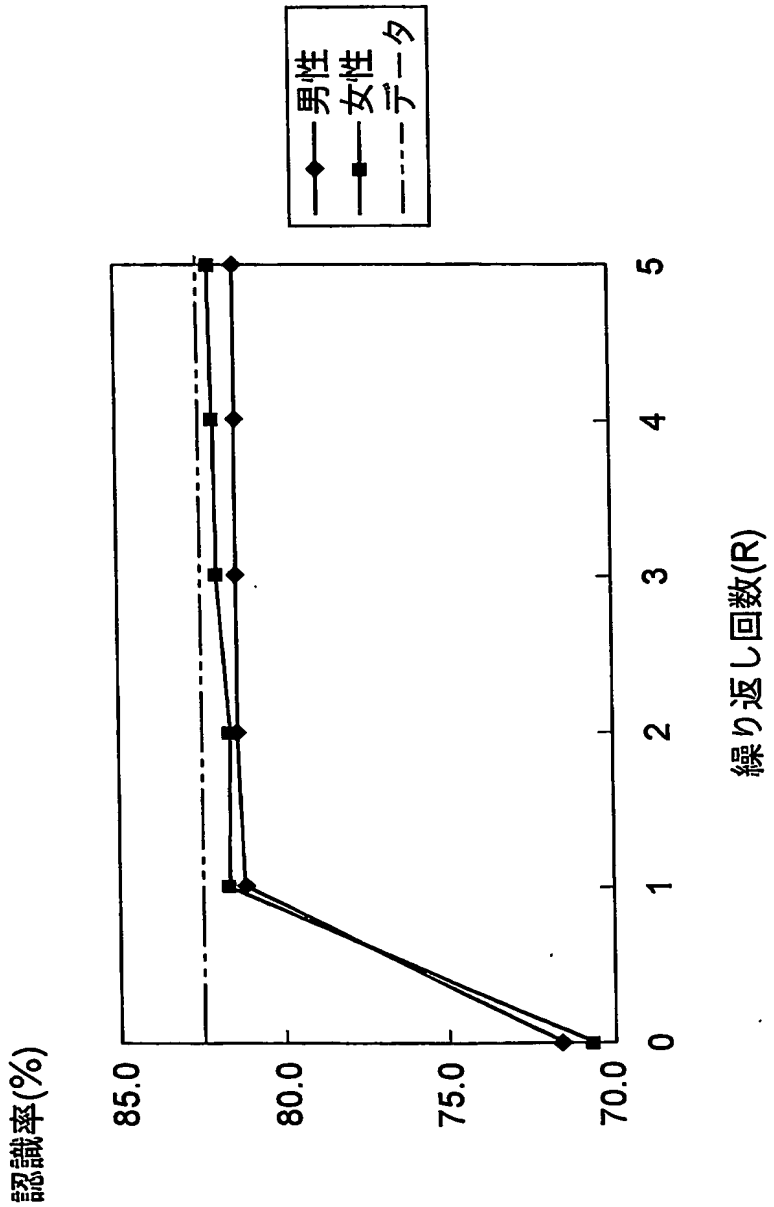


図40

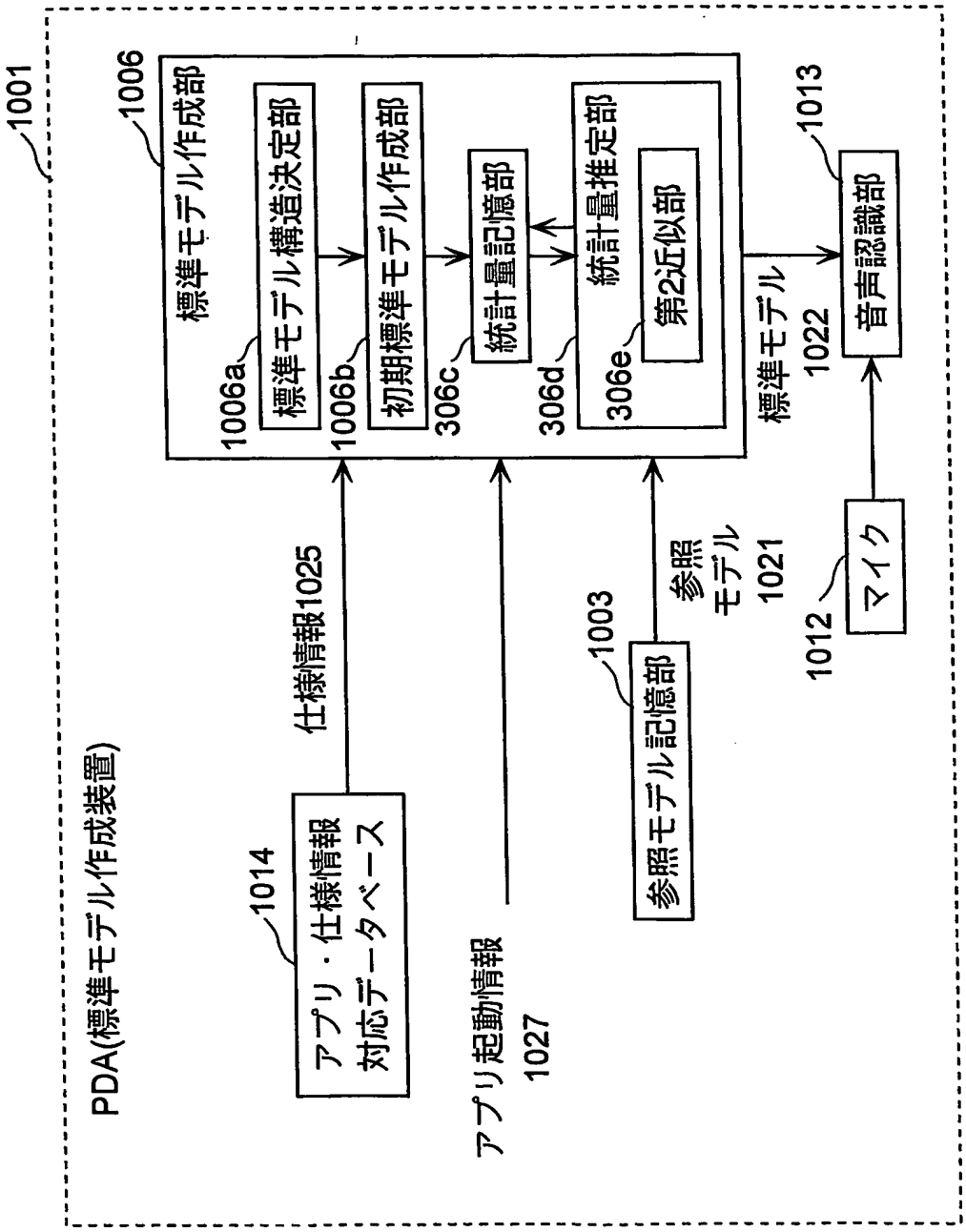

1001
PDA(標準モデル作成装置)
1006
標準モデル作成部
1006a
標準モデル構造決定部
1006b
初期標準モデル作成部
306c
統計量記憶部
306d
統計量推定部
306e
第2近似部
標準モデル
1022
1013
音声認識部
1012
マイク
参照
モデル
1021
1003
参照モデル記憶部
仕様情報1025
1014
アプリ・仕様情報
対応データベース
アプリ起動情報
1027

図41

アプリケーション		仕様情報
ID	名前	
1	ゲームA	混合分布数3
2	ゲームB	混合分布数5
3	株取引	混合分布数126
4	テレビのリモコン	混合分布数5
5	翻訳	混合分布数64

図42

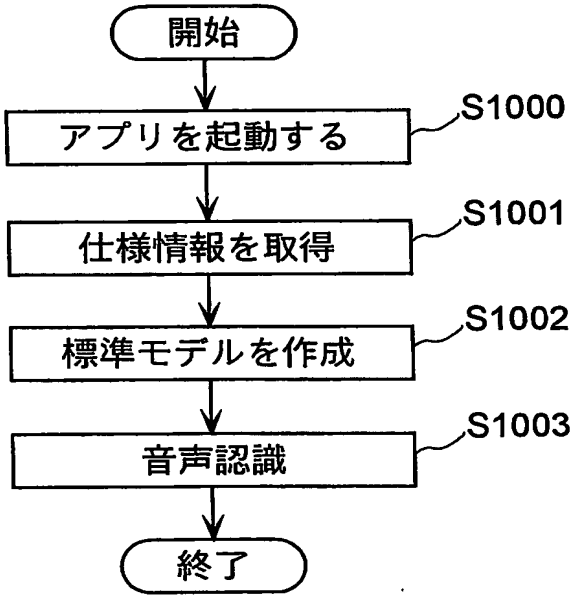

開始
アプリを起動する
S1000
仕様情報を取得
S1001
標準モデルを作成
S1002
音声認識
S1003
終了

図43

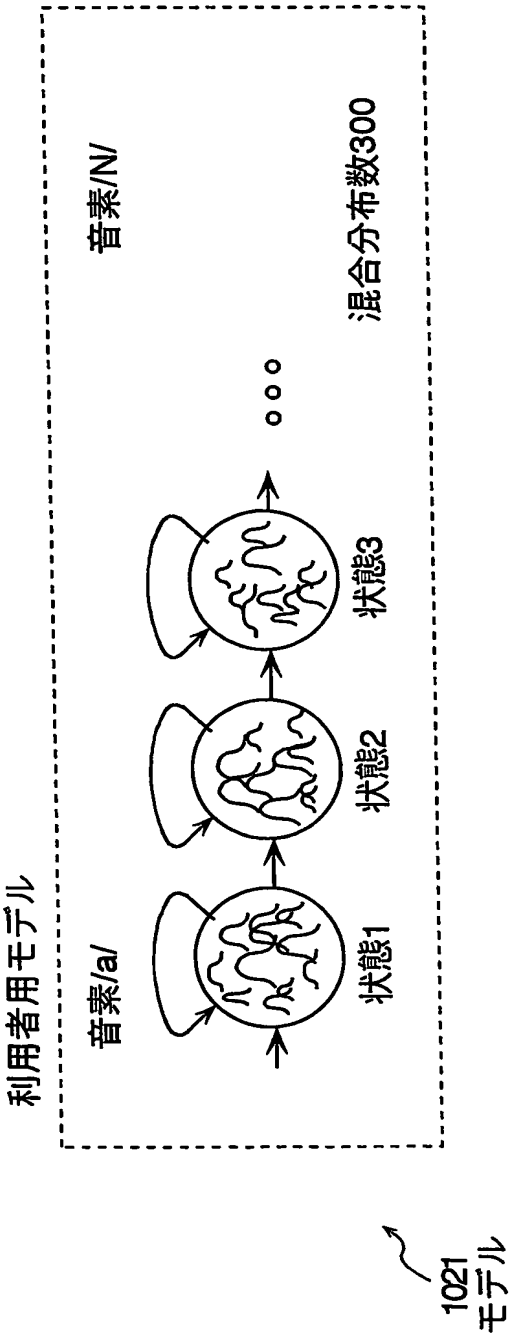


図44

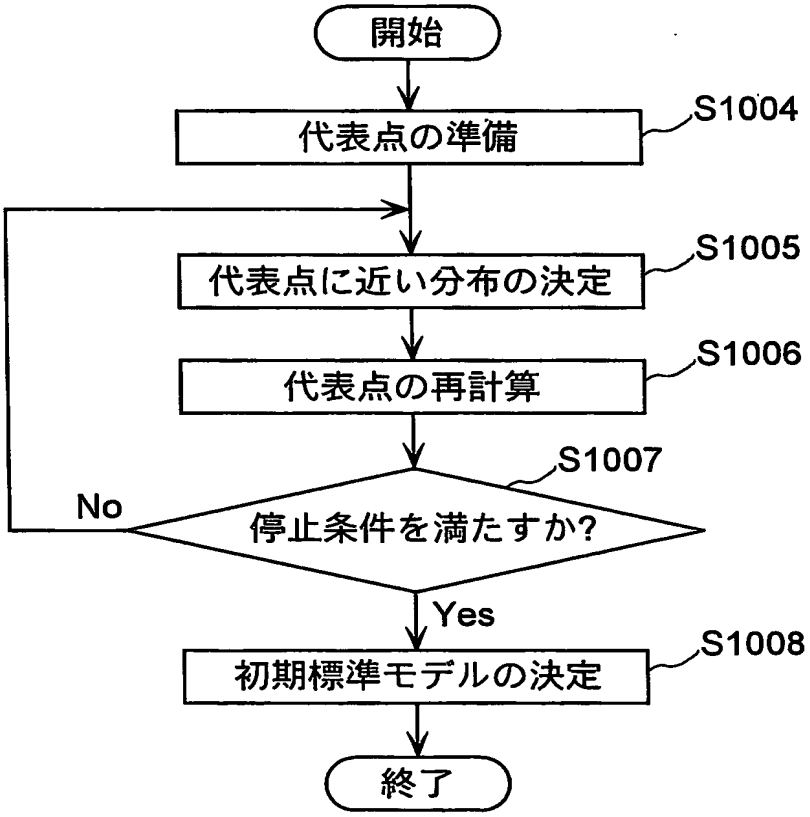

開始
代表点の準備
S1004
代表点に近い分布の決定
S1005
代表点の再計算
S1006
S1007
No
停止条件を満たすか?
Yes
初期標準モデルの決定
S1008
終了

図45

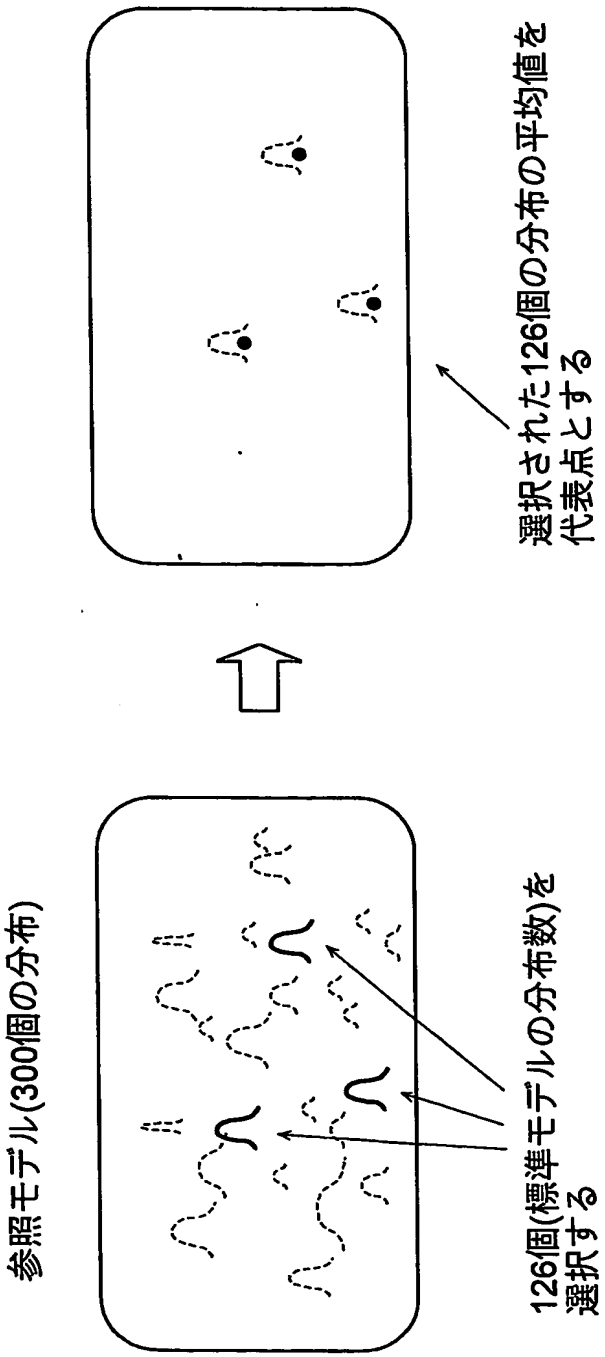

参照モデル(300個の分布)
126個(標準モデルの分布数)を選択する
選択された126個の分布の平均値を代表点とする

図46

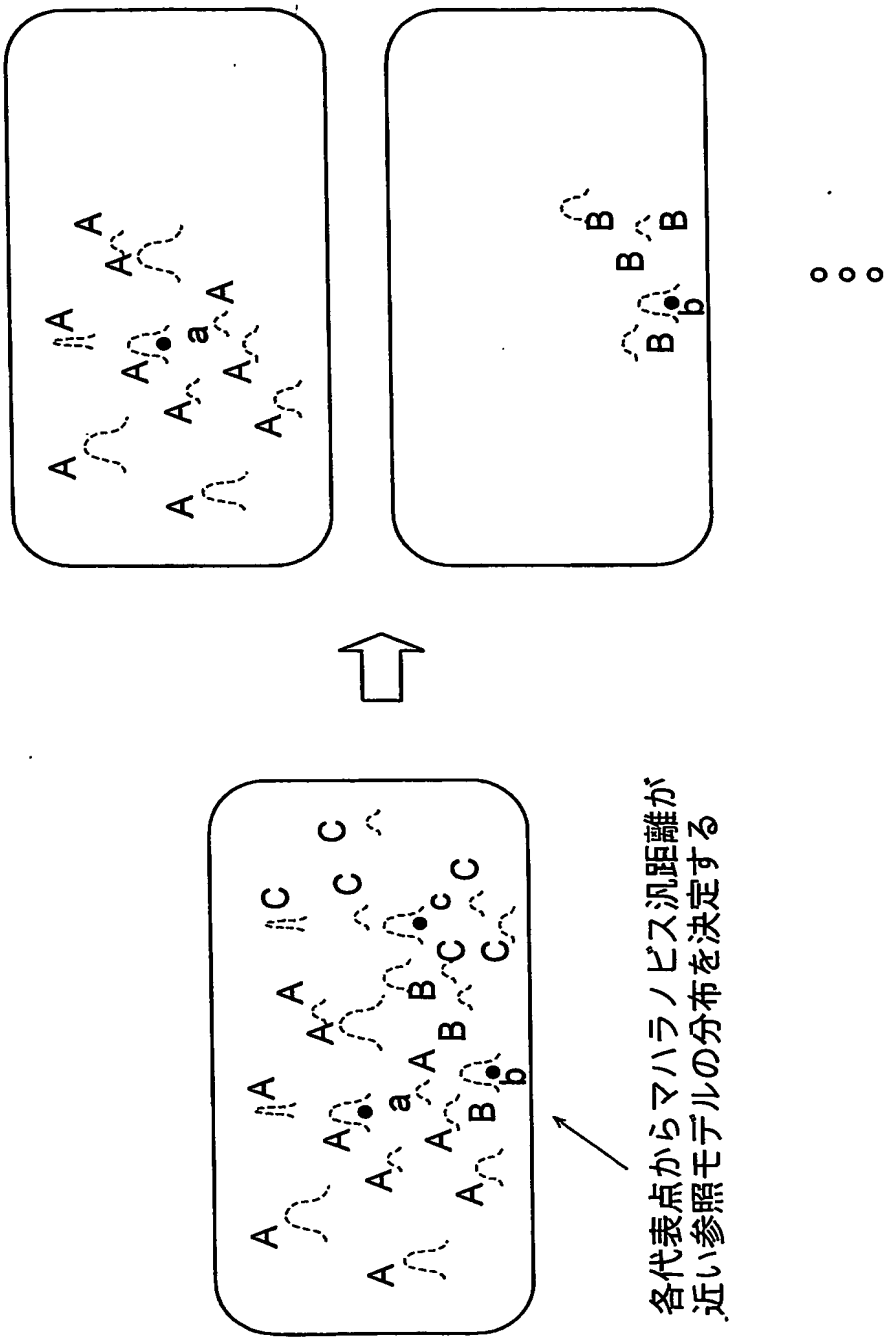


図47

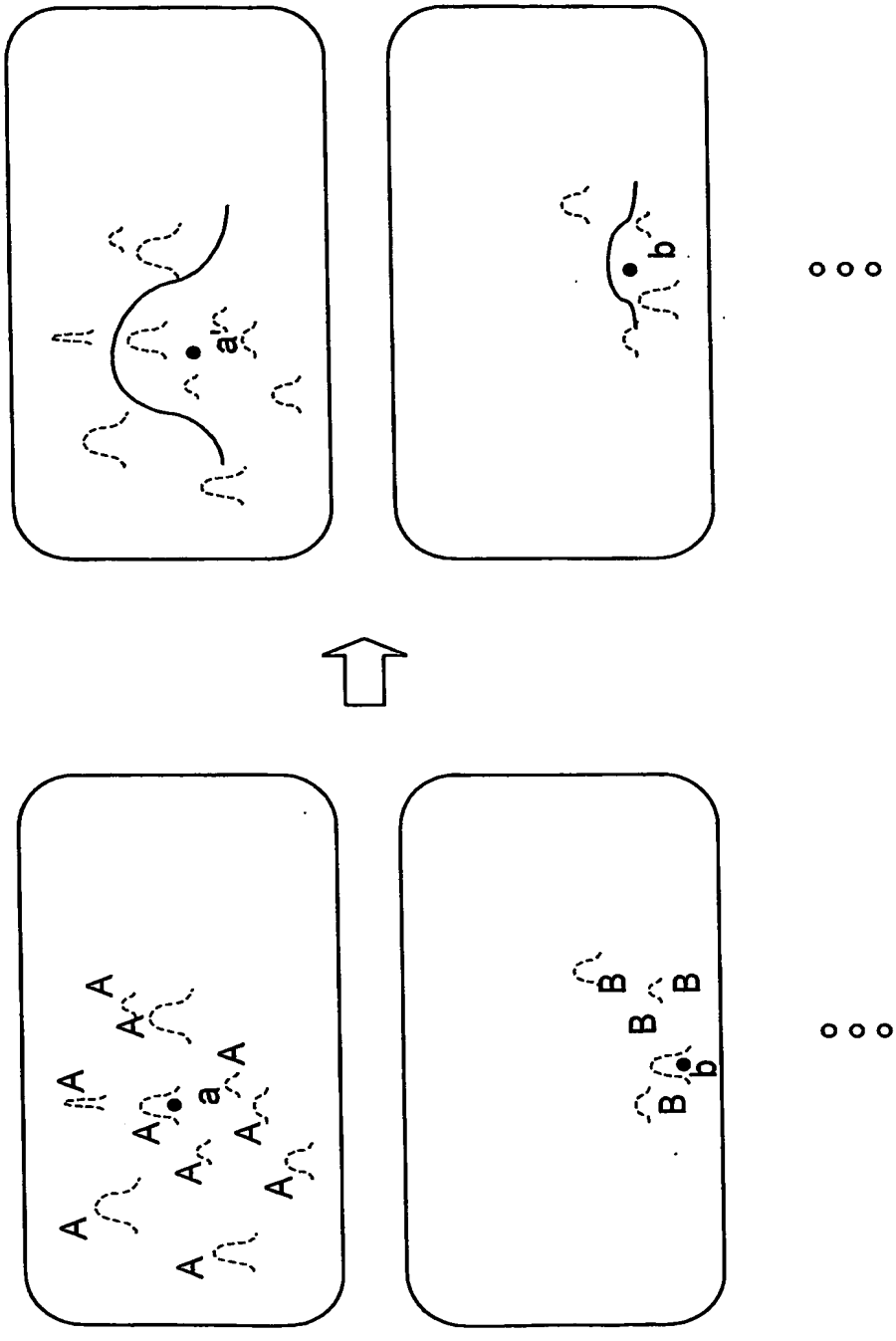


図48

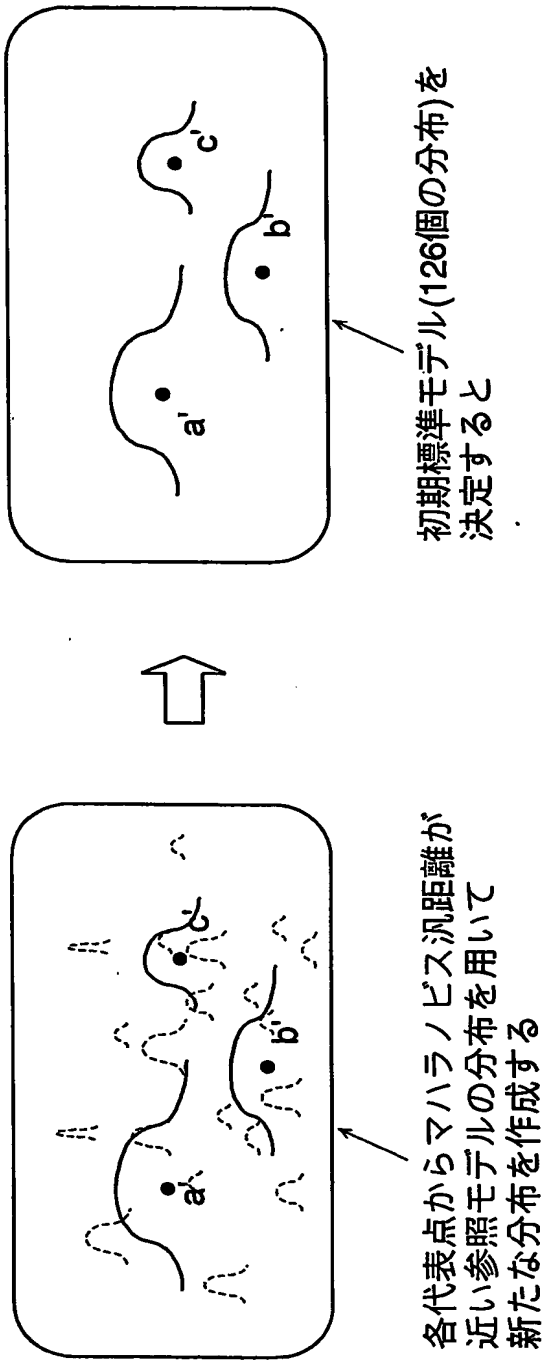

a'
b'
c'
初期標準モデル(126個の分布)を
決定すると
各代表点からマハラノビス汎距離が
近い参照モデルの分布を用いて
新たな分布を作成する

図49

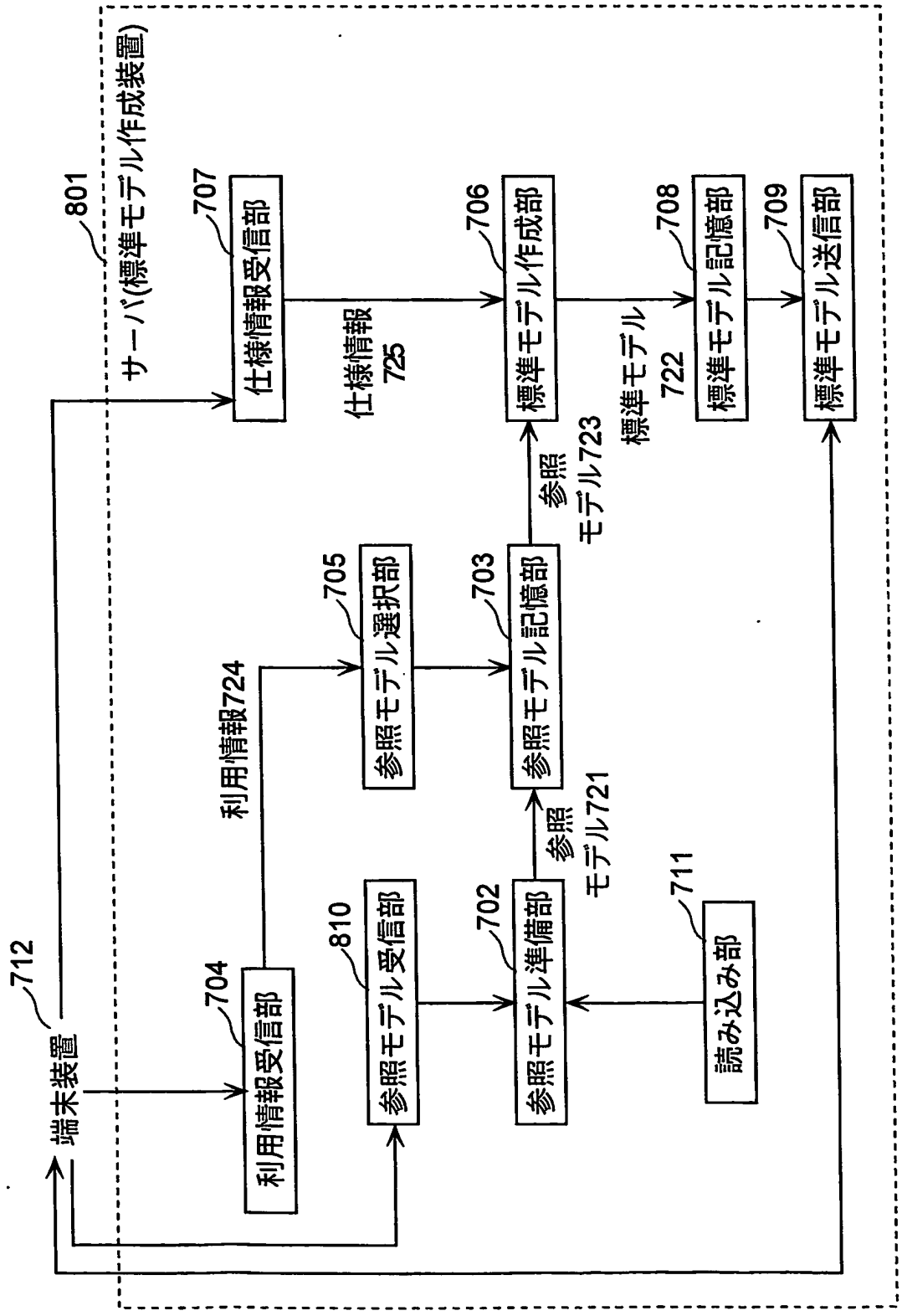

801
サーバ(標準モデル作成装置)
707
仕様情報受信部
706
標準モデル作成部
708
標準モデル記憶部
709
標準モデル送信部
仕様情報
725
標準モデル
722
参照
モデル723
705
参照モデル選択部
703
参照モデル記憶部
利用情報724
712
端末装置
704
利用情報受信部
810
参照モデル受信部
702
参照モデル準備部
参照
モデル721
711
読み込み部

図50

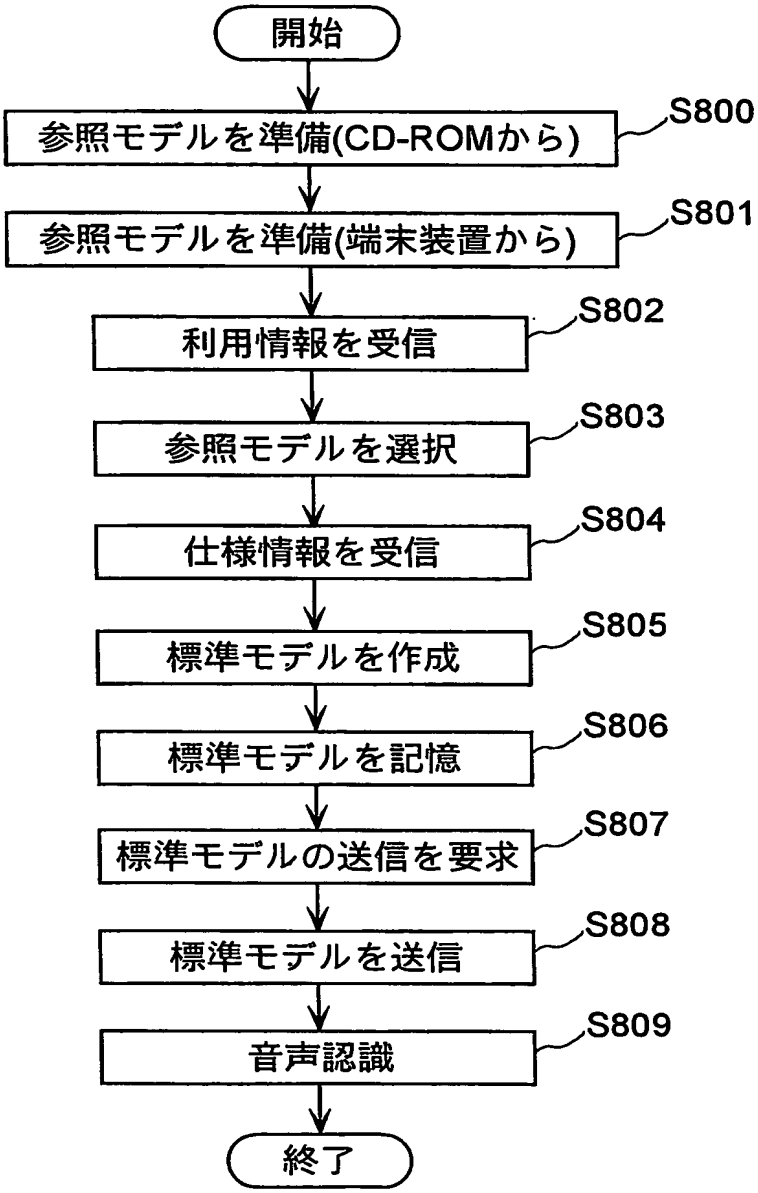

開始
参照モデルを準備(CD-ROMから)
S800
参照モデルを準備(端末装置から)
S801
利用情報を受信
S802
参照モデルを選択
S803
仕様情報を受信
S804
標準モデルを作成
S805
標準モデルを記憶
S806
標準モデルの送信を要求
S807
標準モデルを送信
S808
音声認識
S809
終了

図51

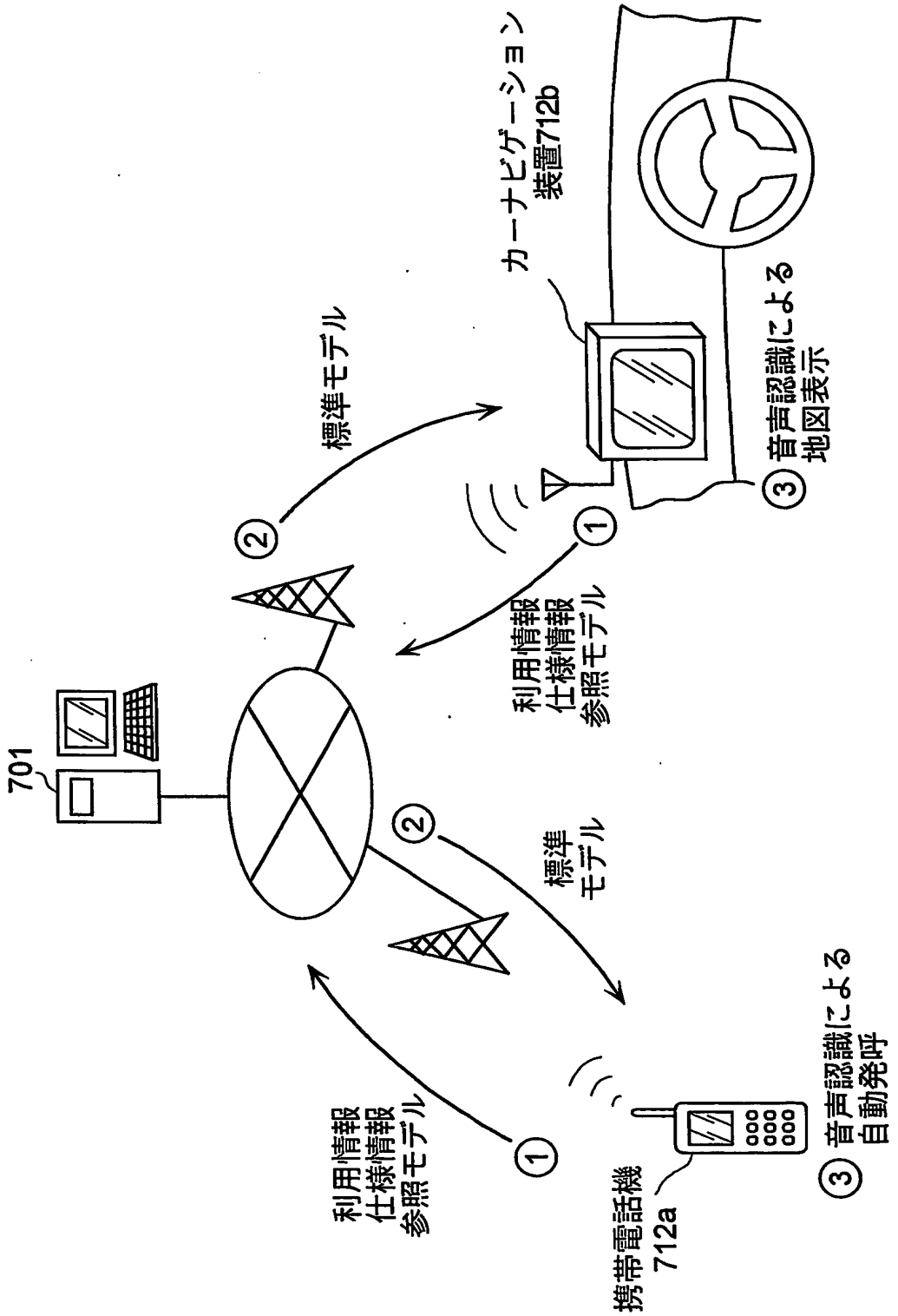

カーナビゲーション
装置712b
標準モデル
②
①
利用情報
仕様情報
参照モデル
③ 音声認識による
地図表示
701
②
標準
モデル
利用情報
仕様情報
参照モデル
①
携帯電話機
712a
③ 音声認識による
自動発呼

図52

クラスID・初期標準モデル・参照モデル対応表

クラスID	初期標準モデル	参照モデル
8A	初期標準モデル8A	参照モデル8AA
		参照モデル8AB
		参照モデル8AC
		⋮
		参照モデル8AZ
⋮	⋮	⋮
64Z	初期標準モデル64Z	参照モデル64ZA
		参照モデル64ZB
		参照モデル64ZC
		⋮
		参照モデル64ZZ

図53

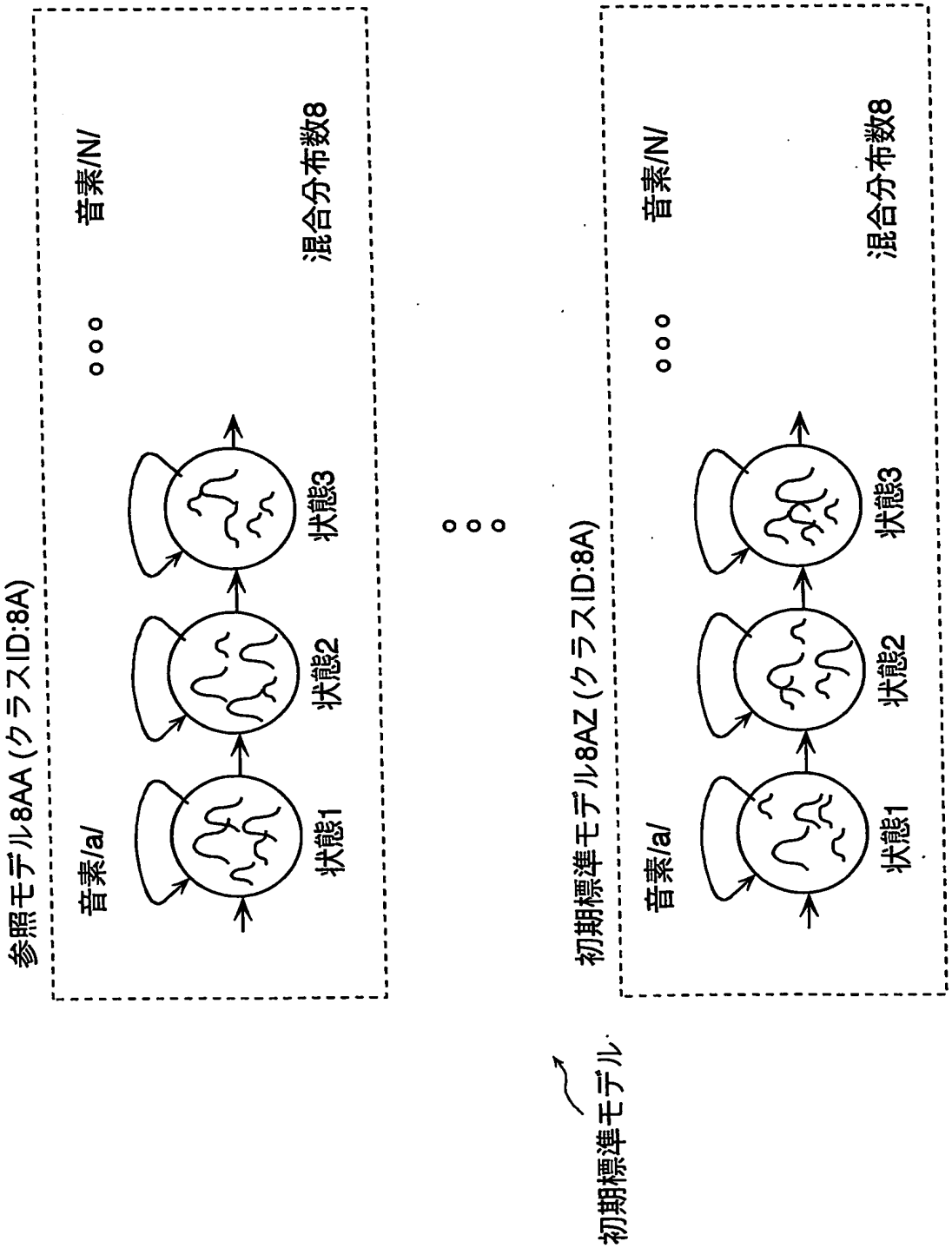

参照モデル8AA (クラスID:8A)
音素/a/
状態1
状態2
状態3
音素/N/
混合分布数8
初期標準モデル8AZ (クラスID:8A)
音素/a/
状態1
状態2
状態3
音素/N/
混合分布数8
初期標準モデル

図54

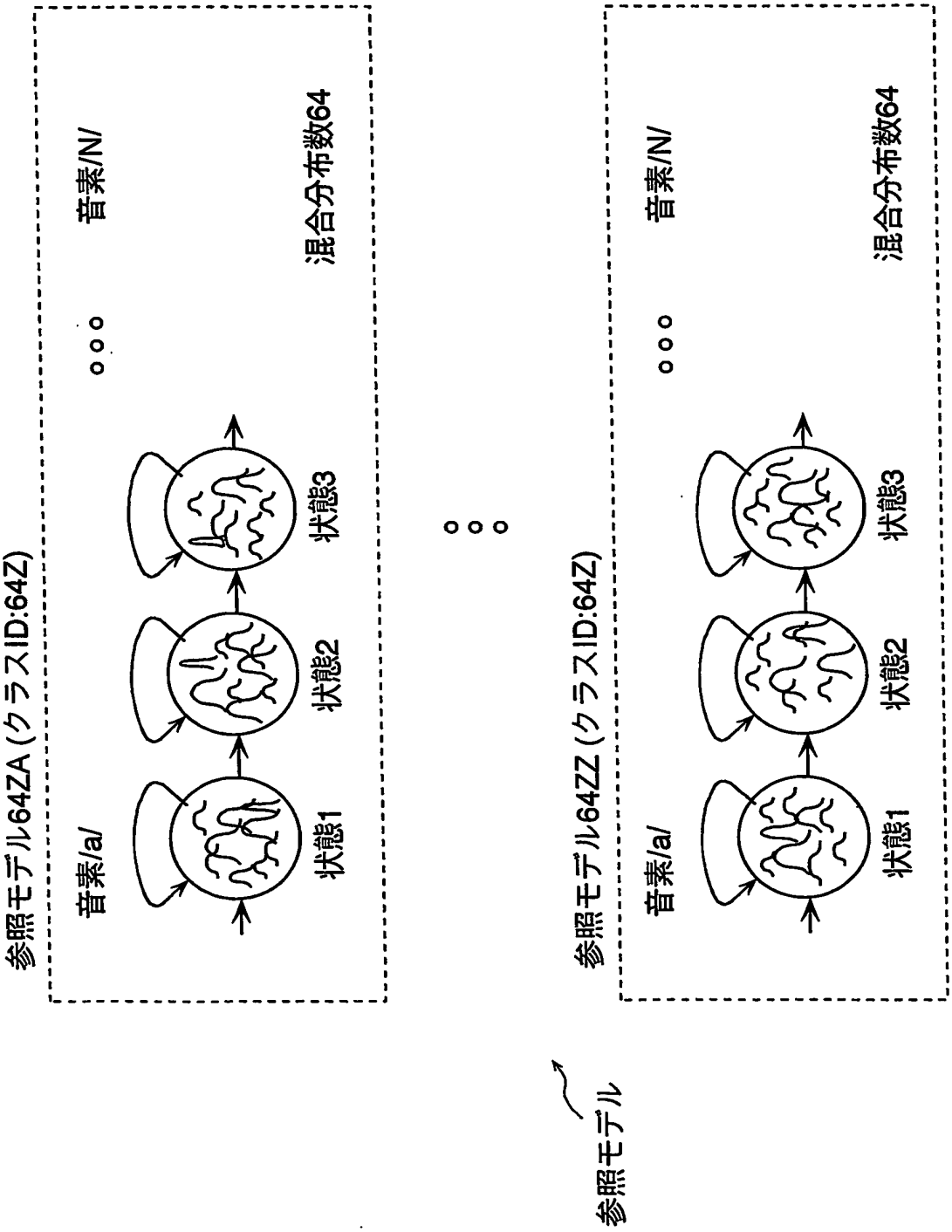

参照モデル
参照モデル64ZA (クラスID:64Z)
音素/a/
状態1
状態2
状態3
音素/N/
混合分布数64
参照モデル64ZZ (クラスID:64Z)
音素/a/
状態1
状態2
状態3
音素/N/
混合分布数64

図55

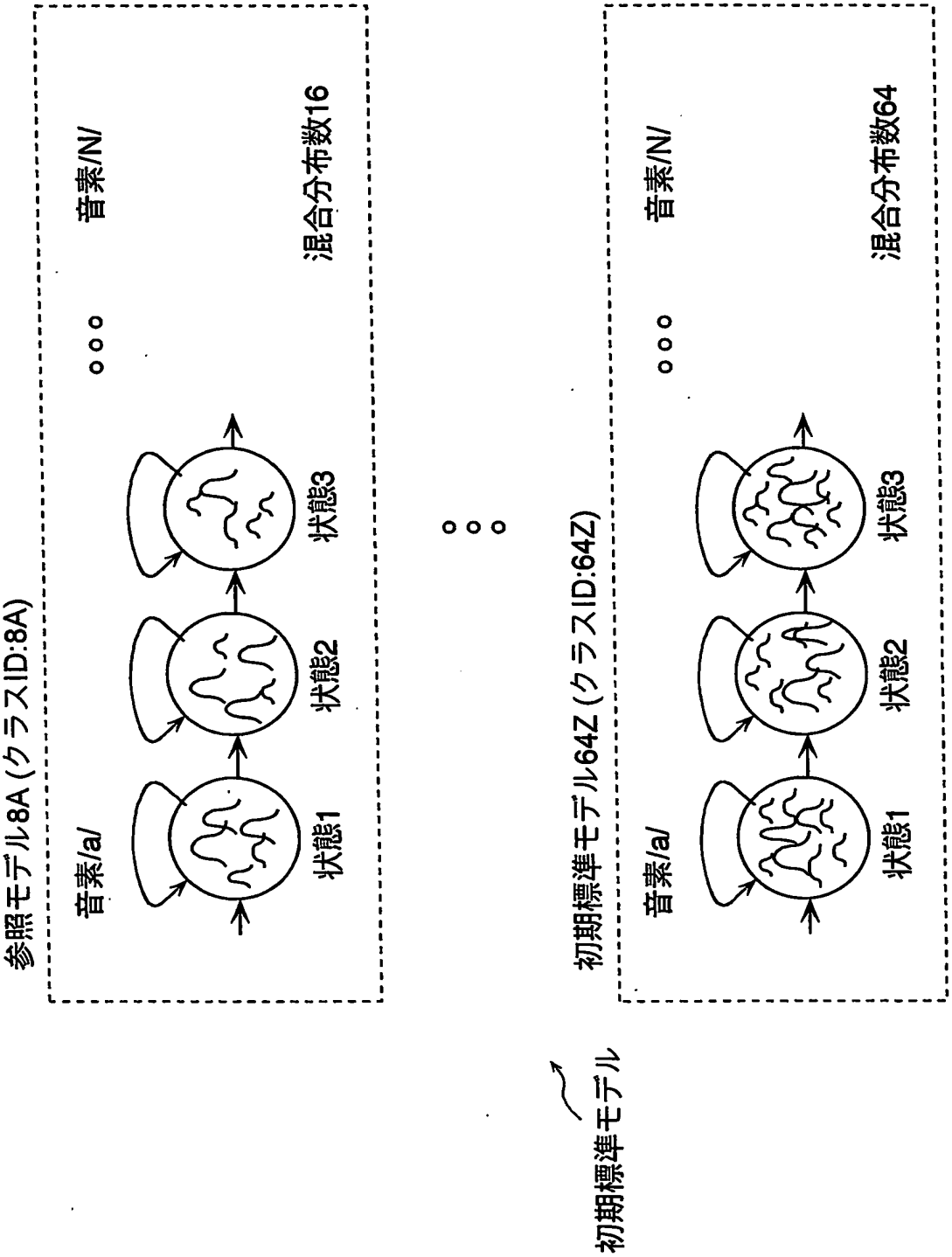

参照モデル8A (クラスID:8A)
音素/a/
音素/N/
状態1
状態2
状態3
混合分布数16
初期標準モデル
初期標準モデル64Z (クラスID:64Z)
音素/a/
音素/N/
状態1
状態2
状態3
混合分布数64

図56

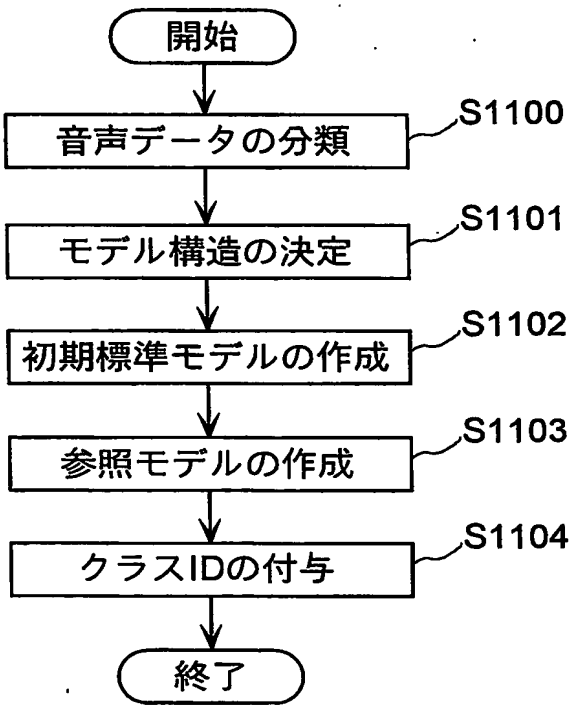

開始
音声データの分類
S1100
モデル構造の決定
S1101
初期標準モデルの作成
S1102
参照モデルの作成
S1103
クラスIDの付与
S1104
終了

図57

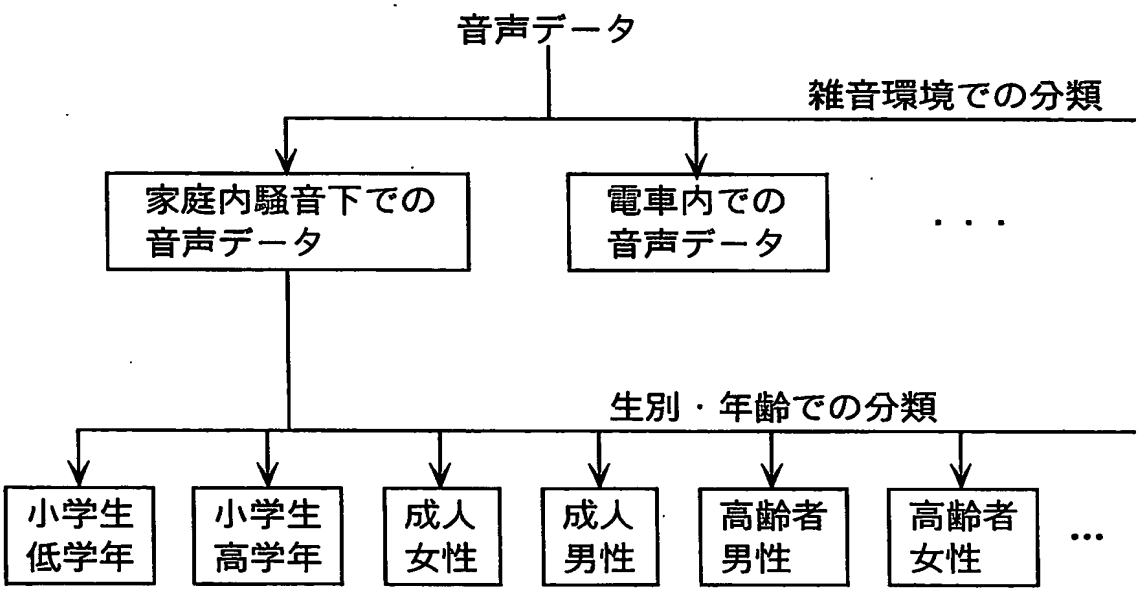

音声データ
雑音環境での分類
家庭内騒音下での
音声データ
電車内での
音声データ
・・・
生別・年齢での分類
小学生
低学年
小学生
高学年
成人
女性
成人
男性
高齢者
男性
高齢者
女性
…

図58

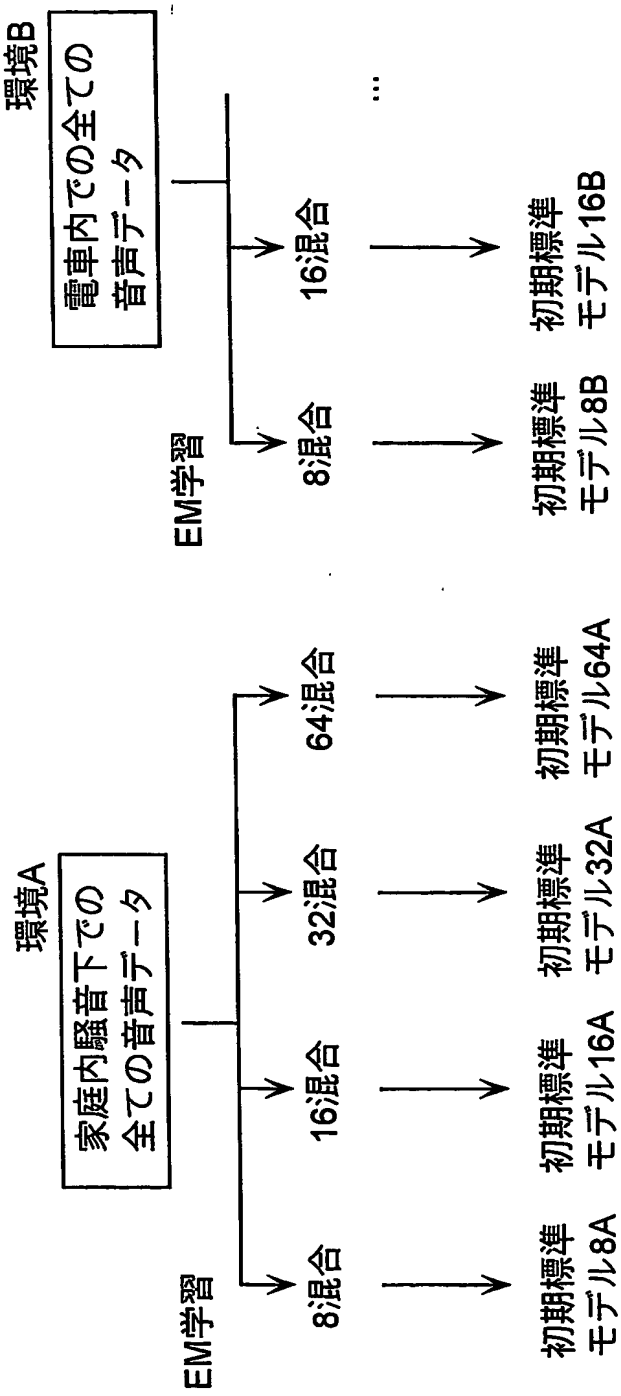

環境A
家庭内騒音下での
全ての音声データ
EM学習
8混合
16混合
32混合
64混合
初期標準
モデル8A
初期標準
モデル16A
初期標準
モデル32A
初期標準
モデル64A
環境B
電車内での全ての
音声データ
EM学習
8混合
16混合
初期標準
モデル8B
初期標準
モデル16B

図59

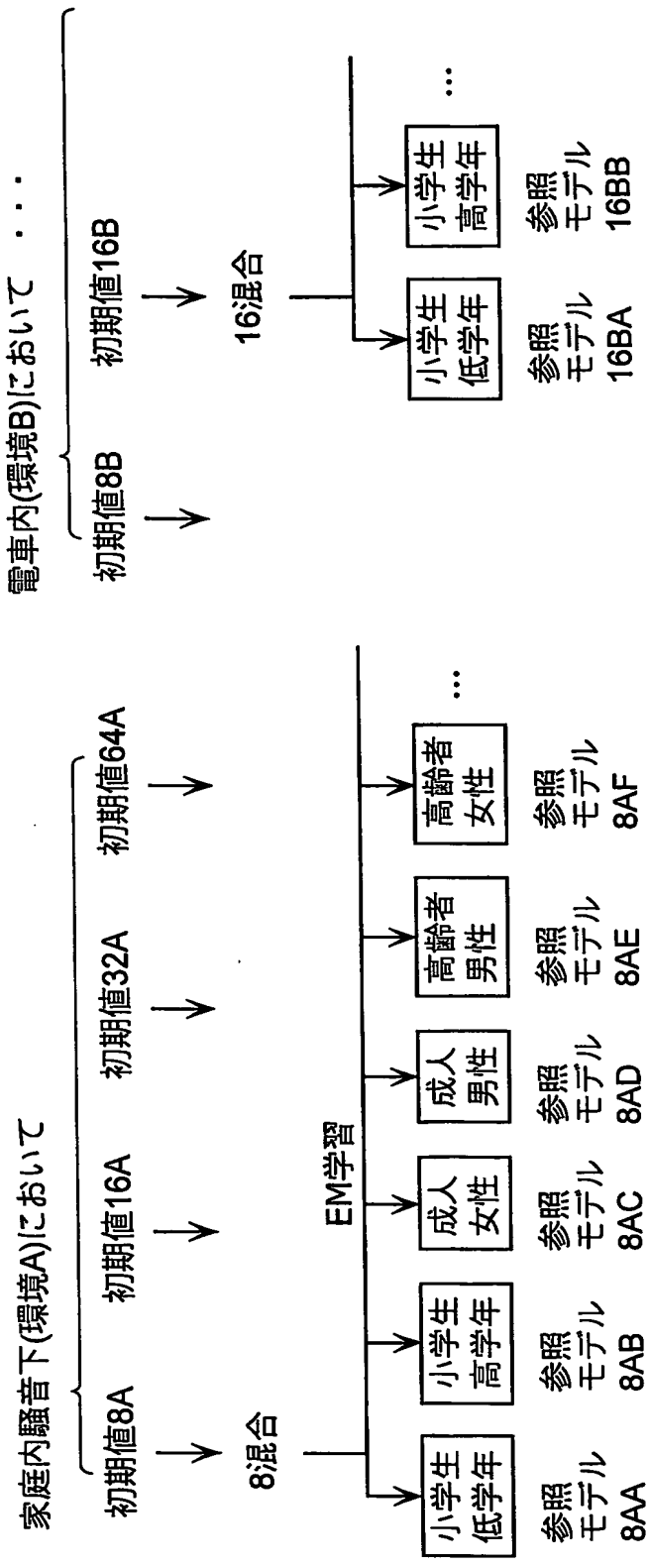

家庭内騒音下(環境A)において
初期値8A
初期値16A
初期値32A
初期値64A
8混合
EM学習
小学生 低学年
参照 モデル 8AA
小学生 高学年
参照 モデル 8AB
成人 女性
参照 モデル 8AC
成人 男性
参照 モデル 8AD
高齢者 男性
参照 モデル 8AE
高齢者 女性
参照 モデル 8AF
…
電車内(環境B)において ・・・
初期値8B
初期値16B
16混合
小学生 低学年
参照 モデル 16BA
小学生 高学年
参照 モデル 16BB
…

図60

クラスID	初期標準モデル	参照モデル	備考(参照モデルの特長)
8A	初期標準モデル8A	参照モデル8AA 参照モデル8AB 参照モデル8AC :	家庭内騒音, 8混合, 小学校低学年 家庭内騒音, 8混合, 小学校高学年 家庭内騒音, 8混合, 成人女性 :
16A	初期標準モデル16A	参照モデル16AA 参照モデル16AB 参照モデル16AC :	家庭内騒音, 16混合, 小学校低学年 家庭内騒音, 16混合, 小学校高学年 家庭内騒音, 16混合, 成人女性 :
:	:	:	:
64B	初期標準モデル64B	参照モデル64BA 参照モデル64BB 参照モデル64BC :	電車内, 64混合, 小学校低学年 電車内, 64混合, 小学校高学年 電車内, 64混合, 成人女性 :

図61

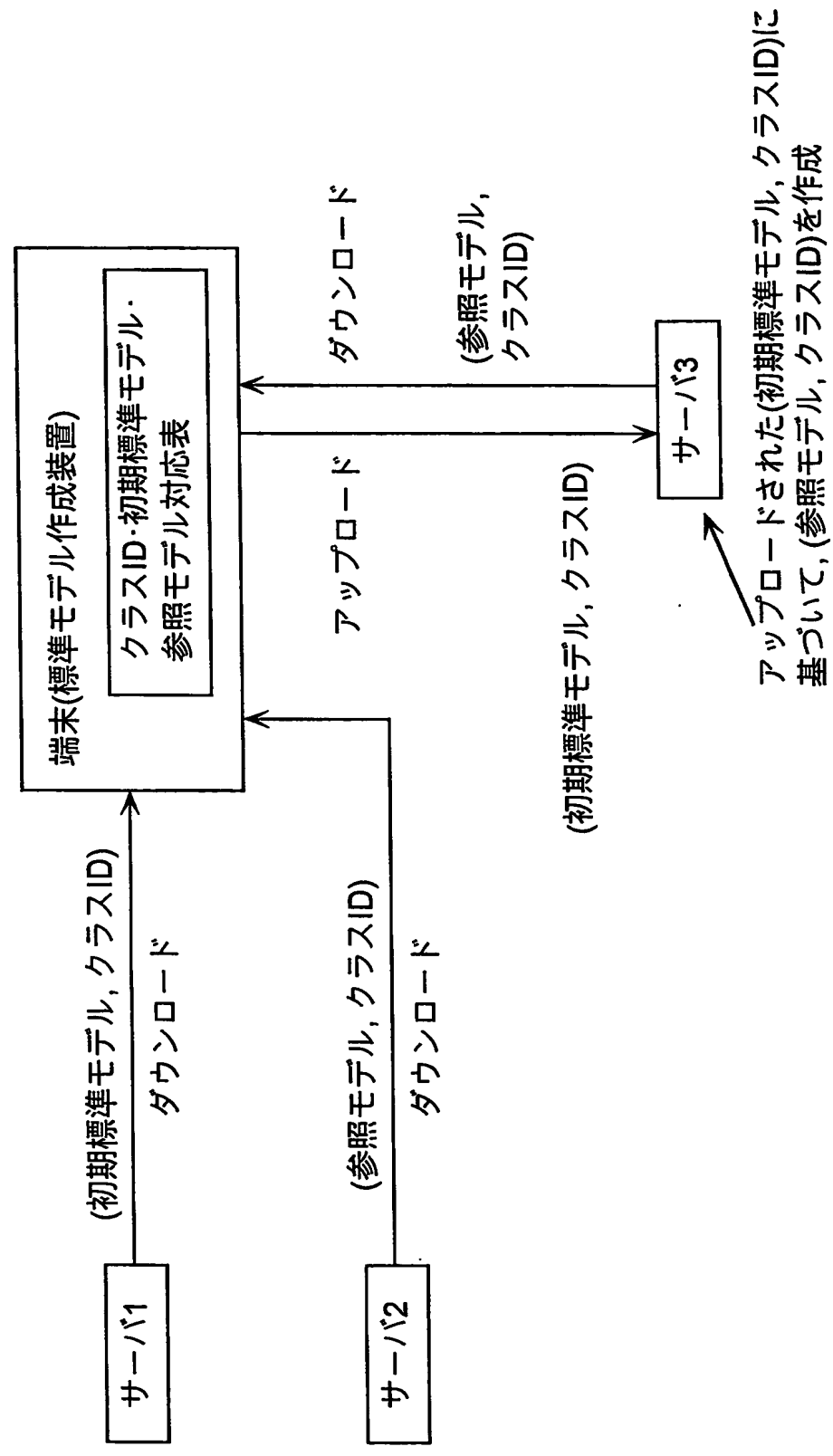

端末(標準モデル作成装置)
クラスID・初期標準モデル・参照モデル対応表
サーバ1
サーバ2
サーバ3
(初期標準モデル, クラスID)
ダウンロード
(参照モデル, クラスID)
ダウンロード
アップロード
(初期標準モデル, クラスID)
ダウンロード
(参照モデル, クラスID)
アップロードされた(初期標準モデル, クラスID)に基づいて, (参照モデル, クラスID)を作成

図62

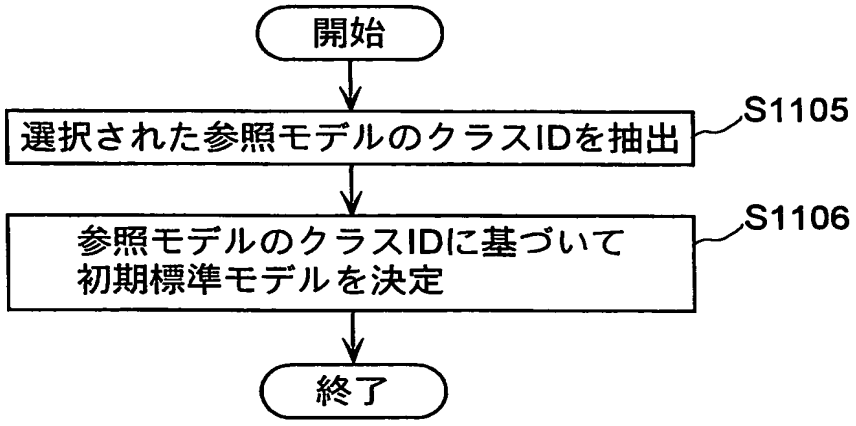

開始
選択された参照モデルのクラスIDを抽出
S1105
参照モデルのクラスIDに基づいて
初期標準モデルを決定
S1106
終了

図63

選択された参照モデル	クラスID
参照モデル8AA	8A
参照モデル16AA	16A
参照モデル16AB	16A
参照モデル16AC	16A
参照モデル16BA	16B
参照モデル64BA	64B

図64

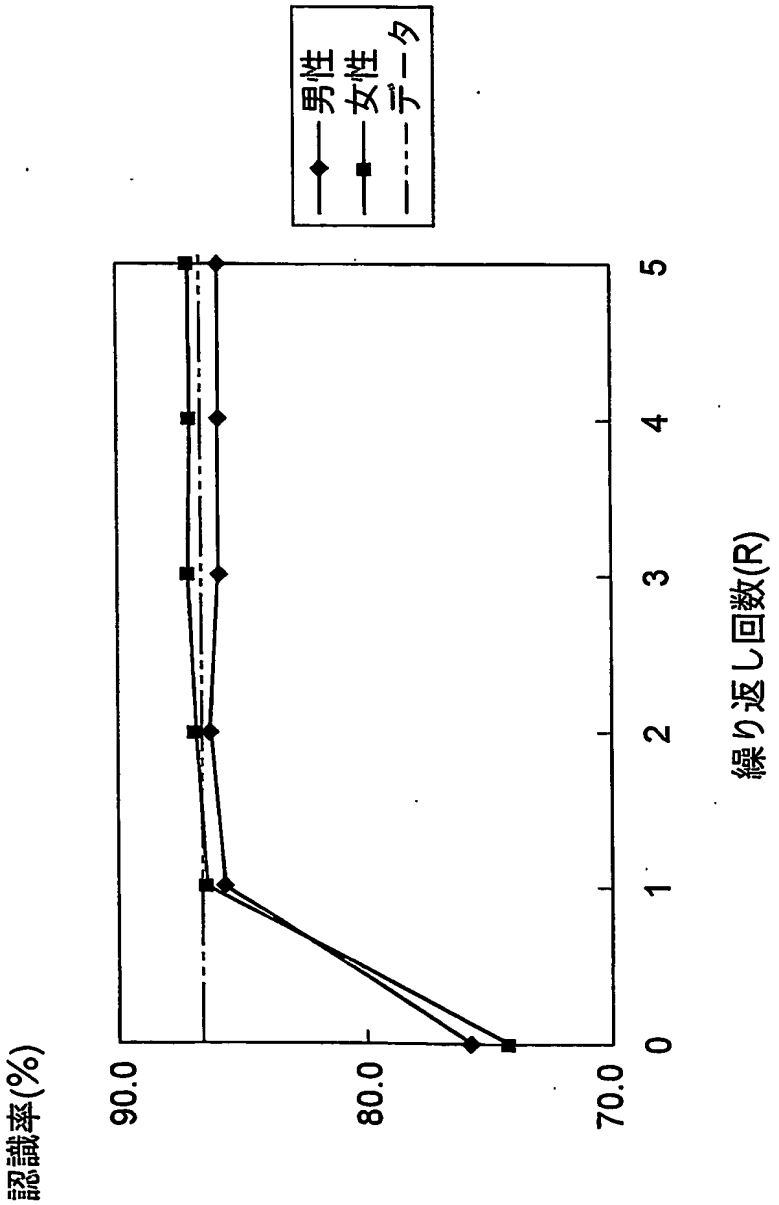
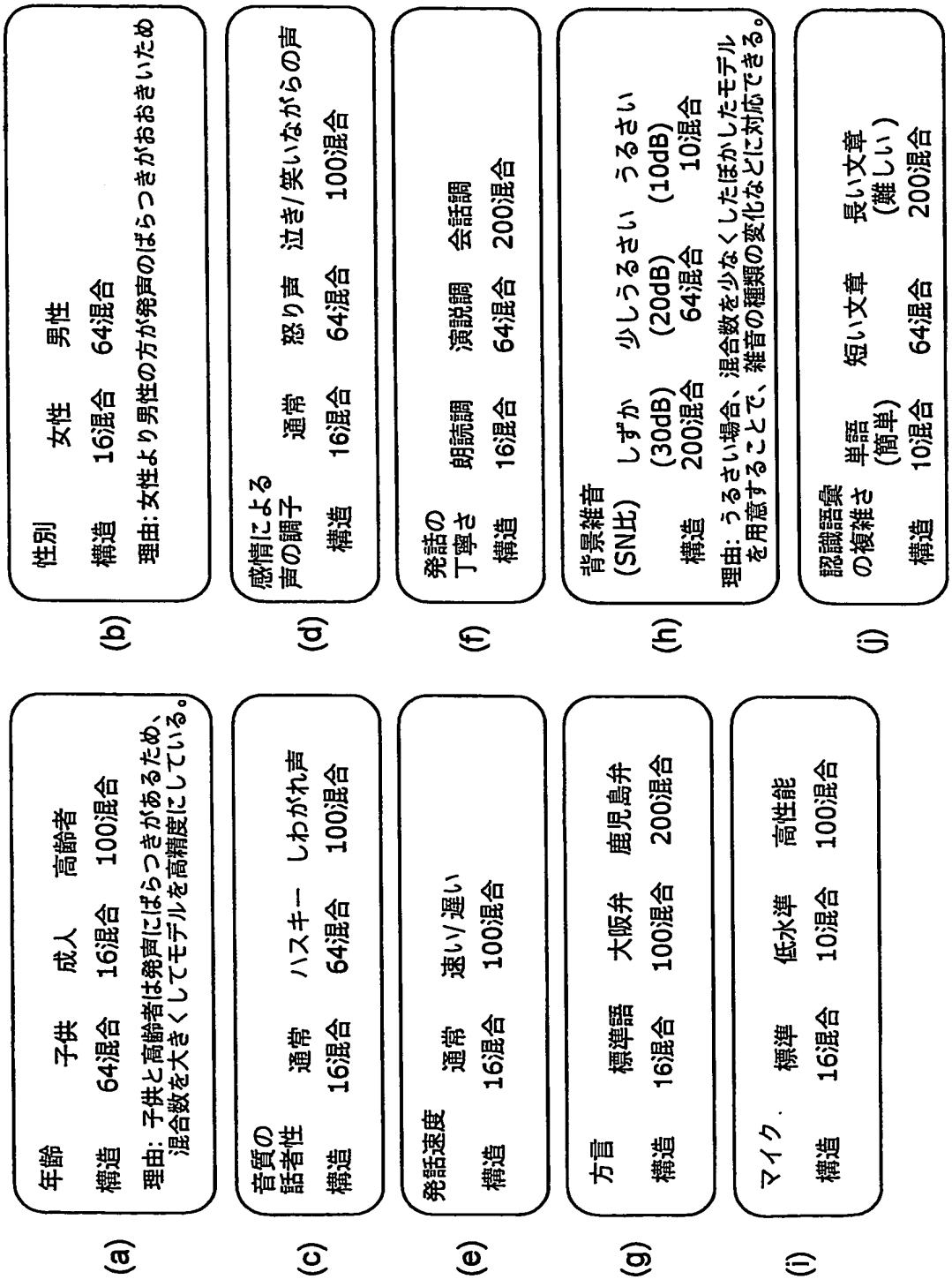


図65


(a)
年齢 子供 成人 高齢者
構造 64混合 16混合 100混合
理由: 子供と高齢者は発声にばらつきがあるため、混合数を大きくしてモデルを高精度にしている。
(b)
性別 女性 男性
構造 16混合 64混合
理由: 女性より男性の方が発声のばらつきがおおきいため
(c)
音質の話者性 通常 ハスキー しわがれ声
構造 16混合 64混合 100混合
(d)
感情による声の調子 通常 怒り声 泣き/笑いながらの声
構造 16混合 64混合 100混合
(e)
発話速度 通常 速い/遅い
構造 16混合 100混合
(f)
発話の丁寧さ 朗読調 演説調 会話調
構造 16混合 64混合 200混合
(g)
方言 標準語 大阪弁 鹿児島弁
構造 16混合 100混合 200混合
(h)
背景雑音(SN比) しずか (30dB) 少しうるさい (20dB) うるさい (10dB)
構造 200混合 64混合 10混合
理由: うるさい場合、混合数を少なくしたぼかしたモデルを用意することで、雑音の種類の変化などに対応できる。
(i)
マイク 標準 低水準 高性能
構造 16混合 10混合 100混合
(j)
認識語彙の複雑さ 単語(簡単) 短い文章 長い文章(難しい)
構造 10混合 64混合 200混合

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP03/14626

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl[7] G10L15/06, G06K9/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int.Cl[7] G10L15/06, G06K9/68

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Jitsuyo Shinan Koho 1926-1995 Toroku Jitsuyo Shinan Koho 1994-2004
Kokai Jitsuyo Shinan Koho 1971-2004 Jitsuyo Shinan Toroku Koho 1996-2004

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
JICST FILE(JOIS), IEEE Xplore

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	YOSHIZAWA, KANO, "Saiyu Suitei ni Motozuku Model Togo Gakushuho", The Acoustical Society of Japan (ASJ) Gakkai 2003 Nen Shuki Kenkyu Happyokai Koen Ronbunshu I, 17 September, 2003 (17.09.03), 3-6-2, pages 105 to 106	1-25
A	YOSHIZAWA et al., "Unsupervised speaker adaptation based on sufficient HMM statistics of selected speakers", Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'01), 07 May, 2001 (07.05.01), Vol.1, pages 341 to 344	1-25
A	JP 11-143486 A (Fuji Xerox Co., Ltd.), 28 May, 1999 (28.05.99), Full text; all drawings (Family: none)	1-25

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
06 January, 2004 (06.01.04)	20 January, 2004 (20.01.04)
Name and mailing address of the ISA/ Japanese Patent Office	Authorized officer
Facsimile No.	Telephone No.

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP03/14626

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 3251005 B2 (NEC Corp.), 16 November, 2001 (16.11.01), Full text; all drawings (Family: none)	1-25
A	JP 7-69711 B2 (Kabushiki Kaisha ATR Jido Honyaku Denwa Kenkyusho), 31 July, 1995 (31.07.95), Full text; all drawings (Family: none)	1-25
A	JP 2002-236494 A (Denso Corp.), 23 August, 2002 (23.08.02), Full text; all drawings (Family: none)	1-25

国際調査報告

国際出願番号 PCT／JP03／14626

A. 発明の属する分野の分類（国際特許分類（IPC））

Int. Cl' G10L15／06, G06K9／68

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

Int. Cl' G10L15／06, G06K9／68

最小限資料以外の資料で調査を行った分野に含まれるもの
日本国実用新案公報 1926～1995年
日本国公開実用新案公報 1971～2004年
日本国登録実用新案公報 1994～2004年
日本国実用新案登録公報 1996～2004年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

JICSTファイル（JOIS）, IEEE Xplore

C. 関連すると認められる文献

引用文献のカテゴリー＊	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求の範囲の番号
PX	芳澤, 鹿野, 「最尤推定に基づくモデル統合学習法」, 日本音響学会2003年秋季研究発表会講演論文集 I, 2003.09.17, 3-6-2, Pages 105-106	1－25
A	Yoshizawa et al, "Unsupervised speaker adaptation based on sufficient HMM statistics of selected speakers", Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '01), 2001.05.07, Volume 1, Pages 341-344	1－25

☒ C欄の続きにも文献が列挙されている。 ☐ パテントファミリーに関する別紙を参照。

＊ 引用文献のカテゴリー
「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
06.01.04	20.1.2004

国際調査機関の名称及びあて先
日本国特許庁（ISA／JP）
郵便番号100－8915
東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員） 5C 9379
樫本 剛
電話番号 03－3581－1101 内線 3541

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ０３／１４６２６

Ｃ（続き）．　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求の範囲の番号
Ａ	ＪＰ　１１－１４３４８６　Ａ（富士ゼロックス株式会社）1999.05.28，全文，全図（ファミリーなし）	１－２５
Ａ	ＪＰ　３２５１００５　Ｂ２（日本電気株式会社）2001.11.16，全文，全図（ファミリーなし）	１－２５
Ａ	ＪＰ　７－６９７１１　Ｂ２（株式会社エイ・ティ・アール自動翻訳電話研究所）1995.07.31，全文，全図（ファミリーなし）	１－２５
Ａ	ＪＰ　２００２－２３６４９４　Ａ（株式会社デンソー）2002.08.23，全文，全図（ファミリーなし）	１－２５